

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:

March 18, 2003, 15:29:05 ; Search time 40 Seconds

(without alignments)
2864.888 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRGSGYPHLLMDVKKRSLG.....RLKEDRSGSGQENENDEDE 860

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched:

908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:

908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 120 summaries

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	4514	99.7	22	Human Repro-EN-1.0
2	4498	99.4	22	Human cell cycle a
3	4455.5	98.6	22	Human IBI protein.
4	2390	52.8	22	Human signal trans
5	1459	32.2	22	Human diagnostic a
6	847	18.7	22	Human protein sequ
7	668	15.2	22	Human secreted pro
8	664	14.7	22	Human secreted pro
9	569	12.6	22	Peptide #2607 enco
10	369	12.6	22	Protein #2554 enco

11	569	12.6	110	22	AA55960	Human brain expres
12	569	12.6	110	22	AA68328	Human bone marrow
13	569	12.6	110	22	AA61614	Peptide #2578 enco
14	569	12.6	110	22	AA03874	Peptide #2556 enco
15	569	12.6	110	22	AB37880	Human peptide enco
16	508.5	11.2	628	22	AB61380	Drosophila melanog
17	508.5	11.2	628	22	AA080462	Fruit fly adipose
18	499.5	11.0	677	23	AA080464	Human adipose prot
19	489.5	10.6	677	23	AA080466	Human adipose prot
20	481.5	10.6	536	23	AA080470	Fruit fly adipose
21	470.5	10.4	580	23	AA080471	Human adipose prot
22	460.5	10.2	580	23	AA080472	Mouse adipose prot
23	441	9.7	616	23	AA080473	Human adipose prot
24	437	9.7	86	22	AB27840	Human adipose prot
25	437	9.7	86	22	AB33011	Human adipose prot
26	437	9.7	86	22	AB18484	Peptide #517 enco
27	437	9.7	86	22	AA03809	Protein #483 enco
28	437	9.7	86	22	AA03809	Human bone marrow
29	437	9.7	86	22	AA03809	Peptide #509 enco
30	437	9.7	86	22	AA03809	Peptide #488 enco
31	437	9.7	86	22	AA03809	Human peptide enco
32	437	9.7	86	22	AA03809	Arabidopsis thalia
33	436.5	9.6	471	21	AA05041	Arabidopsis thalia
34	436.5	9.6	471	21	AA05041	Novel human diagno
35	429.5	8.9	816	16	AA05041	Novel human diagno
36	404	8.9	633	22	AB27224	Novel human diagno
37	399	8.8	385	23	AB080461	Fruit fly adipose
38	373	8.2	748	22	AB080461	Drosophila melanog
39	338	7.5	62	22	AB080461	Human peptide #492
40	338	7.5	62	22	AB080461	Human peptide #518 enco
41	338	7.5	62	22	AB080461	Protein #484 enco
42	338	7.5	62	22	AB080461	Human brain expres
43	338	7.5	62	22	AB080461	Human bone marrow
44	338	7.5	62	22	AB080461	Peptide #501 enco
45	338	7.5	62	22	AB080461	Peptide #489 enco
46	338	7.5	62	22	AB080461	Human peptide enco
47	338	7.5	62	22	AB080461	Human peptide enco
48	330	7.3	576	23	AA080472	Human adipose prot
49	329	6.8	60	22	AA080472	Human Repro-EN-1.0
50	329	6.8	60	22	AA080472	Human Repro-EN-1.0
51	296	6.5	60	22	AA080472	Human prostate can
52	273	6.0	214	21	AA080472	Arabidopsis thalia
53	269.5	6.0	365	21	AA080472	Arabidopsis thalia
54	269.5	6.0	365	21	AA080472	Human protein sequ
55	264	5.8	139	21	AA080472	Human protein sequ
56	245.5	5.4	355	22	AA080472	Human protein sequ
57	244.5	5.3	355	22	AA080472	Human protein sequ
58	238.5	5.3	358	21	AA080472	Arabidopsis thalia
59	238.5	5.3	358	21	AA080472	Arabidopsis thalia
60	234.5	5.2	191	21	AA080472	Arabidopsis thalia
61	234.5	5.2	191	21	AA080472	Arabidopsis thalia
62	234	5.2	212	21	AA080472	Arabidopsis thalia
63	232.5	5.1	942	22	AA080472	Human Repro-EN-1.0
64	230	5.1	174	21	AA080472	Human immune/thaema
65	214	4.7	37	22	AA080472	Human polypeptide
66	213.5	4.7	37	22	AA080472	Human polypeptide
67	205.5	4.5	747	22	AA080472	Human polypeptide
68	204.5	4.5	1301	22	AA080472	Human polypeptide
69	204.5	4.5	1219	22	AA080472	Human polypeptide
70	203	4.5	1219	22	AA080472	Human polypeptide
71	201	4.4	2137	23	AA080472	Human polypeptide
72	200	4.4	2137	23	AA080472	Human polypeptide
73	200	4.4	2137	23	AA080472	Human polypeptide
74	195.5	4.3	2858	22	AA080472	Human polypeptide
75	194.5	4.3	2858	22	AA080472	Human polypeptide
76	194.5	4.3	2858	22	AA080472	Human polypeptide
77	193	4.3	3060	22	AA080472	Human polypeptide
78	192.5	4.3	985	16	AA080472	Human polypeptide
79	192.5	4.3	985	16	AA080472	Human polypeptide
80	191.5	4.2	2951	22	AA080472	Human polypeptide
81	191	4.2	2951	22	AA080472	Human polypeptide
82	189	4.2	1508	21	AA080472	Human polypeptide
83	189	4.2	1508	21	AA080472	Human polypeptide
84	189	4.2	1508	21	AA080472	Human polypeptide
85	189	4.2	1508	21	AA080472	Human polypeptide
86	189	4.2	1508	21	AA080472	Human polypeptide
87	189	4.2	1508	21	AA080472	Human polypeptide
88	189	4.2	1508	21	AA080472	Human polypeptide
89	189	4.2	1508	21	AA080472	Human polypeptide
90	189	4.2	1508	21	AA080472	Human polypeptide
91	189	4.2	1508	21	AA080472	Human polypeptide
92	189	4.2	1508	21	AA080472	Human polypeptide
93	189	4.2	1508	21	AA080472	Human polypeptide
94	189	4.2	1508	21	AA080472	Human polypeptide
95	189	4.2	1508	21	AA080472	Human polypeptide
96	189	4.2	1508	21	AA080472	Human polypeptide
97	189	4.2	1508	21	AA080472	Human polypeptide
98	189	4.2	1508	21	AA080472	Human polypeptide
99	189	4.2	1508	21	AA080472	Human polypeptide
100	189	4.2	1508	21	AA080472	Human polypeptide

Amilo acid sequenc
Novel human diago
Novel human diago
Human polypeptide
Amilo acid sequenc
Novel human diago
Novel OREF OREF1857
Streptococcus pen
S. pneumoniae SP00
Human OREF OREF2780
Drosophila melanog
Novel human diago
Staphylococcus aur
Human zc1 protein.
Novel human diago
Human HEP/GSK-1like
Human kinase (PRIN
Human human diago
Novel human diago
Plasmodium falcipa
Human RepO-EN-1.0
Peptide #2335 enco
Peptide #2403 enco
protein #2313 enco
Human brain expres
Human bone marrow
Peptide #3327 enco
Human peptide enco
Lactococcus lactis
Drosophila melanog
Novel human diago
Staphylococcus aur
Staphylococcus aur
Mouse GREO2-relate
Human OREF OREF2889
Drosophila melanog
Human dentin matril

RESULT 1
AAB69556 standard; Protein; 860 AA.

27-APR-2001 (first entry)

Human Repro-EN-1.0 protein.

Human; Repro-EN-1.0; IBI; cytostatic; breast cancer; uterine cancer; prostate cancer; epitope.

~~Homo sapiens~~

WO200107616-A1.

01-FEB-2001.

2000WO-US06742.

1000

30-JUL-1999

23-NOV-57

(DIAG-) DIA

El Shami AS

WPI; 2001-1

N-PSDB, 1998

1

XX New autoantigens Repro-EN-1.0 and IB 1 polypeptides are useful for pathologic
PT are useful for diagnosing endometriosis or as a marker for pathologic
PT conditions such as breast, uterine or prostate cancer

XX Claim 23; Page 22-27; 119pp; English.

PS The present sequence is given in a specification relating to recombinant
XX polynucleotides comprising nucleotide sequences encoding a polypeptide
CC polynucleotides comprising nucleotide sequences encoding a polypeptide
CC epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a
CC fully defined sequence of 860 and 937 amino acids. The epitope
CC specifically binds to antibodies from subjects diagnosed with
CC endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for
CC diagnosing endometriosis by detecting antibodies in immunoassays, and
CC are used as markers for pathologic conditions e.g. breast, uterine or
CC prostatic cancer. Methods for detecting Repro-EN-1.0 or IB 1
CC polynucleotides or polypeptides are useful in the diagnosis of these
CC cancers, monitoring their progress or treatment, and determining patient
CC prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1
CC may be used as probes for detecting mRNA from cell types suspected to
CC be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and
CC IB 1 polypeptides and immunogenic fragments may be used as positive
CC controls in diagnostic assays to detect antibodies that specifically
CC bind the proteins from patient serum samples. The polypeptides are
CC useful as immunogens for eliciting antibody production against epitopes
of the protein, and as controls in diagnostic methods.

XX Sequence 860 AA;

XX

4714; DB 22; Length 860;

Query Match	99.78%	Pred. No. 0;	
Best Local Similarity	99.78%		
Matches	857;	Mismatches	1;
	Conservative		Indels
			Gaps
			0;

[illegible]

00	1	DTGEYILSGDDTKLVI	NSPYRKKVLTTR	SGHRANIFSAKEL	PCYNDKQIV	SCSGDGYI	120
01	2	61	DTGEYILSGDDTKLVI	NSPYRKKVLTTR	SGHRANIFSAKEL	PCYNDKQIV	SCSGDGYI 120

Db 61 DTGEYILSGSDDFKLVISNPYSRKVLTIRSGHRANLFSANLDFCNDNG-----

	EVTNNEODAEINROCO.....	121	FVTNNEODAEINROCO.....	180
OY				
	EVTNNEODAEINROCOFTCHYGTYEIMNYPNDBYTFLSCGEDGVKRFDRIKTSCKE	131	EVTNNEODAEINROCOFTCHYGTYEIMNYPNDBYTFLSCGEDGVKRFDRIKTSCKE	190

181 DCKDDILNCRATSVACPPIPYLAVGCDSSVRIYDRMLGTRATGNAGKGIIGM 240

Db 181 DKKDILINCRRAATSAVACIPILPIILAVGGSDGDAK...
...ATTTTTCGCGGCTGCGAGGGGGLVYSSDYITLEDPKDDTARELKTPEAERRE 300

241 VAREIPSHLNKSCRAFTSLCYSEDGDEILVATSSDIYLFDPKDTARELKTSPSEERRE 300

0Y ELRQPPVKRLRLRGDMSDTGPRAPRPESEERERQGEQSPNVSIMQKMSDMUUNVZ 360

Db 301 ELRQPVKRLKLDMDJG: 420

[illegible]

421 AHSTSPESHSTPLSSPDSEQRQVEASGHHHNGD.....
|||||
|||TCTCMEPQPSVFAAGHTTHQSDNNNEKLSPKPGTGEPVLSLH 480
|||||

QY

Db 422 AHS155FLD:.....
A81 VSTEGTTTGIKLPDMSIASSRGISGCHCKSECEESTVPOSSVQPPGDSSEKAP 540

```

Db 481 YSTEGITTTIKLNFIDEMSSIASSSRNGISGHCKSEQEESEFVQFFEDDB.....

```

[illegible]

—

—

QY 601 SSEKAEPTSDOTSEATNTNNTNPEPQFOTEGPSAHEETSTRDALSODTDDSD 660
 Db 601 SSEKAEPTSDOTSEATNTNNTNPEPQFOTEGPSAHEETSTRDALSODTDDSD 660
 QY 661 PVLIGARYRAGPDRSARVARIOEFERRRERKEEMEDLTINIRPLVKMYKGRNR 720
 Db 661 PVLIGARYRAGPDRSARVARIOEFERRRERKEEMEDLTINIRPLVKMYKGRNR 720
 QY 721 TMIKANTWGANFVWNSGDCGHIFTWDRHTAHEMLLEADNHNVCLOPPEPDIASSG 780
 Db 721 TMIKANTWGANFVWNSGDCGHIFTWDRHTAHEMLLEADNHNVCLOPPEPDIASSG 780
 QY 781 IDYDIKISPLEESRIENRKLADDEVITRNEMLLEETRTITYPASPMLMLASLNHTRAD 840
 Db 781 IDYDIKISPLEESRIENRKLADDEVITRNEMLLEETRTITYPASPMLMLASLNHTRAD 840
 QY 841 RLEGDRSESGOENENEDEE 860
 Db 841 RLEGDRSESGOENENEDEE 860

RESULT 2

AAB60505
 ID AAB60505 standard; Protein: 880 AA.

AC AAB60505;

DT 24-APR-2001 (first entry)

DE Human cell cycle and proliferation protein CCYPR-53, SEQ ID NO:53.

KM Cell cycle and proliferation protein; CCYPR; human; agonist;

KM antagonist; gene therapy; detection; gene therapy;

KM transgenic animal disease model; immune disorder;

KM developmental disorder; cell signalling disorder;

KM cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

KM arteriosclerosis; asthma; allergy; diabetes mellitus;

KM menstrual cycle disorder; bacterial infection.

OS Homo sapiens.

XX WO200107471-A2.

XX 01-FEB-2001.

PE 21-JUL-2000; 2000MO-US19948.

PR 21-JUL-1999; 99US-0145075.

PR 08-SEP-1999; 99US-0153129.

PR 10-NOV-1999; 99US-0164447.

XX (INCY-) INCYTE-GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

XX Azimail Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

XX MPI; 2001-11-27/72.

XX N-PSDB; AAF59642.

XX

PT Human cell cycle and proliferation proteins and polynucleotides are

PT used to treat, diagnose and prevent immune, developmental and cell

PT signalling disorders and cell proliferative disorders including cancer

XX Claim 1; Page 164-166; 205pp; English.

PS

CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and

CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.

CC CCYPR and agonists of CCYPR are used to treat diseases or conditions

CC associated with decreased expression of functional CCYPR, while CCYPR

CC antagonists are used to treat diseases or conditions associated with

CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies

CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or

CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect

CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders, and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC caused by bacteria.

XX Sequence 880 AA;

Query Match 99.4%; Score 4498; DB 22; Length 880;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 859; Conservative 0; Mismatches 1; Indels 20; Gaps 1;

QY 1 MSRGSGYPHLLMDVRRKSLGLEDPRLNSRYLGRREFIQRLKLEATLVNHDCCVNTICWN 60
 Db 1 MSRGSGYPHLLMDVRRKSLGLEDPRLNSRYLGRREFIQRLKLEATLVNHDCCVNTICWN 60
 QY 61 DTGEYILSGDDTKLYISNPYSRKVLTTRSGHRANIFSAKFLPCTNKKQIYSCSGDGYI 120
 Db 61 DTGEYILSGDDTKLYISNPYSRKVLTTRSGHRANIFSAKFLPCTNKKQIYSCSGDGYI 120
 QY 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWDFTRIKTSCTRE 180
 Db 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWDFTRIKTSCTRE 180
 QY 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWDFTRIKTSCTRE 180
 Db 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWDFTRIKTSCTRE 180
 QY 181 DCKDDILNCRRAAASVAICPPPIYLLAVGCDSSVRIYDRMLCTRATGAGAGTTGM 240
 Db 181 DCKDDILNCRRAAASVAICPPPIYLLAVGCDSSVRIYDRMLCTRATGAGAGTTGM 240
 QY 241 VARETPSHLNKSKRYTSLCYSEDEQELIVSYSDIYLFDPKDTARLEKTPSAEERE 300
 Db 241 VARETPSHLNKSKRYTSLCYSEDEQELIVSYSDIYLFDPKDTARLEKTPSAEERE 300
 QY 241 VARETPSHLNKSKRYTSLCYSEDEQELIVSYSDIYLFDPKDTARLEKTPSAEERE 300
 Db 241 VARETPSHLNKSKRYTSLCYSEDEQELIVSYSDIYLFDPKDTARLEKTPSAEERE 300
 QY 301 ELROPVYKRLRLRGDMSDTPRAPRESERERDEQSPNVSLMQRSDMLSRFEASAVA 360
 Db 301 ELROPVYKRLRLRGDMSDTPRAPRESERERDEQSPNVSLMQRSDMLSRFEASAVA 360
 QY 361 QSNRGKRSRPRGTSOSDISTLPYVSSPDLEVSETAMEVDTPABOFLQSTSTMSAQ 420
 Db 361 QSNRGKRSRPRGTSOSDISTLPYVSSPDLEVSETAMEVDTPABOFLQSTSTMSAQ 420
 QY 421 AHSTSSPTSPHSPLSSPDSEORQSYEASGHTHHQS-----D 460
 Db 421 AHSTSSPTSPHSPLSSPDSEORQSYEASGHTHHQS-----D 460
 QY 461 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISCHCKSEGOEE 520
 Db 461 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISCHCKSEGOEE 520
 QY 481 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISCHCKSEGOEE 540
 Db 481 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISCHCKSEGOEE 540
 QY 521 SFVPOSSVQPEGSEETAPRESSSDYTKYQEGVSAENPVNHNINOSDKFTAKPLDSN 580
 Db 521 SFVPOSSVQPEGSEETAPRESSSDYTKYQEGVSAENPVNHNINOSDKFTAKPLDSN 580
 QY 541 SFVPOSSVQPEGSEETAPRESSSDYTKYQEGVSAENPVNHNINOSDKFTAKPLDSN 600
 Db 541 SFVPOSSVQPEGSEETAPRESSSDYTKYQEGVSAENPVNHNINOSDKFTAKPLDSN 600
 QY 581 SGERNDLNDRCGVPESASSEKAKBETSDQSTESATNNTNPEPQFOTEGATGSA 640
 Db 581 SGERNDLNDRCGVPESASSEKAKBETSDQSTESATNNTNPEPQFOTEGATGSA 640
 QY 601 SGERNDLNDRCGVPESASSEKAKBETSDQSTESATNNTNPEPQFOTEGATGSA 660
 Db 601 SGERNDLNDRCGVPESASSEKAKBETSDQSTESATNNTNPEPQFOTEGATGSA 660
 QY 641 HEETSTRDALSODTDDSDPVLLPGARYRAGPDRSARVARIOEFERRRERKEEMELD 700
 Db 641 HEETSTRDALSODTDDSDPVLLPGARYRAGPDRSARVARIOEFERRRERKEEMELD 700
 QY 661 HEETSTRDALSODTDDSDPVLLPGARYRAGPDRSARVARIOEFERRRERKEEMELD 720
 Db 661 HEETSTRDALSODTDDSDPVLLPGARYRAGPDRSARVARIOEFERRRERKEEMELD 720
 QY 701 TLTNRPLVAVKVGKSHNSRTMIKEANFWGANFVWNSGDCGHIFTWDRHTAHEMLLEAD 760
 Db 701 TLTNRPLVAVKVGKSHNSRTMIKEANFWGANFVWNSGDCGHIFTWDRHTAHEMLLEAD 760
 QY 721 TLTNRPLVAVKVGKSHNSRTMIKEANFWGANFVWNSGDCGHIFTWDRHTAHEMLLEAD 780
 Db 721 TLTNRPLVAVKVGKSHNSRTMIKEANFWGANFVWNSGDCGHIFTWDRHTAHEMLLEAD 780

QY 761 NHVNCLOPHPPPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNEMLLETRNTI 820
 DB 781 NHVNCLOPHPPPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNEMLLETRNTI 840
 QY 821 TVPASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 860
 DB 841 TVPASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 880
 RESULT 3
 AAB69557
 ID AAB69557 standard; Protein; 937 AA.
 AC AAB69557;
 DT 27-APR-2001 (first entry)
 DE Human IB1 protein.
 XX Human: Repro-EN-1.0; IB1: cytostatic; breast cancer; uterine cancer;
 KW prostate cancer; epitope.
 OS Homo sapiens.
 PN MO200107616-A1.
 PD 01-FEB-2001.
 PF 10-MAR-2000; 2000MO-US06742.
 PR 22-JUL-1999; 99MO-0359084.
 PR 30-JUL-1999; 99MO-US17284.
 PR 23-NOV-1999; 99US-0447399.
 XX (DIAG-) DIAGNOSTIC PROD CORP.
 PI E1 Shaml AS; Menon SN, French CK;
 DR WPI; 2001-182795/18.
 DR N-PSDB; AAF58655.
 PT New autoantigens Repro-EN-1.0 and IB 1 polypeptides and polynucleotides
 PT are useful for diagnosing endometriosis or as a marker for pathologic
 PT conditions such as breast, uterine or prostate cancer
 PS Claim 24; Page 29-34; 119pp; English.
 CC The present sequence is given in a specification relating to recombinant
 CC polynucleotides comprising nucleotide sequences encoding a polypeptide
 CC epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a
 CC fully defined sequence of 860 and 937 amino acids. The epitope
 CC specifically binds to antibodies from subjects diagnosed with
 CC endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for
 CC diagnosing endometriosis by detecting antibodies in immunoassays, and
 CC are used as markers for pathologic conditions e.g. breast, uterine or
 CC prostate cancer. Methods for detecting Repro-EN-1.0 or IB 1
 CC polynucleotides or polypeptides are useful in the diagnosis of these
 CC cancers, monitoring their progress or treatment, and determining patient
 CC prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1
 CC may be used as probes for detecting mRNA from cell types suspected to
 CC be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and
 CC IB 1 polypeptides and immunogenic fragments may be used as positive
 CC controls in diagnostic assays to detect antibodies that specifically
 CC bind the proteins from patient serum samples. The polypeptides are
 CC useful as immunogens for eliciting antibody production against epitopes
 CC of the protein, and as controls in diagnostic methods.
 CC Sequence 937 AA;
 XX
 Query Match 98.6%; Score 4465.5; DB 22; Length 937;
 Best local similarity 91.5%; Pred. No. 3.8e-318;
 Matches 857; Conservative 2; Mismatches 1; Indels 77; Gaps 1;

QY 1 MSRGSGYPHLLMDVRRKSLGLEDEPSRLRSRYLGRREFIQRLLKLEATLNVHDCVNTICWN 60
 DB 1 MSRGSGYPHLLMDVRRKSLGLEDEPSRLRSRYLGRREFIQRLLKLEATLNVHDCVNTICWN 60
 QY 61 DTGEYIISGDDPTKLYISNYSRKVLTITSGHRANIFSAKFLPCTNDKQIVSGSGGY 120
 DB 61 DTGEYIISGDDPTKLYISNYSRKVLTITSGHRANIFSAKFLPCTNDKQIVSGSGGY 120
 QY 121 FYTNVEDAEFTNQCCPTCHGTYYEIMTVPNPDYTFILSCGEDGTVMFPDRIKTSCKE 180
 DB 121 FYTNVEDAEFTNQCCPTCHGTYYEIMTVPNPDYTFILSCGEDGTVMFPDRIKTSCKE 180
 QY 181 DCRKDDILINCRRATSVAILCPPIPYTLAVGSDSSVRIYDRMLGTAAGNAGRTTGM 240
 DB 181 DCRKDDILINCRRATSVAILCPPIPYTLAVGSDSSVRIYDRMLGTAAGNAGRTTGM 240
 QY 241 VAFIPSHLNKSCRYSLCYSEDOGLIVSYSDYLYLPDKDPTARELKTGPAERRE 300
 DB 241 VAFIPSHLNKSCRYSLCYSEDOGLIVSYSDYLYLPDKDPTARELKTGPAERRE 300
 QY 301 ELRQPPVKRLRLRGDWSDTGPRAPESERERDQSPNVSIMQMSDMLSMFPEASEVA 360
 DB 301 ELRQPPVKRLRLRGDWSDTGPRAPESERERDQSPNVSIMQMSDMLSMFPEASEVA 360
 QY 361 QSNRGGRSRPRGTSQSDISTLPTVPSSPDLVSETAMEVDTPAEQFLQSTSTSAQ 420
 DB 361 QSNRGGRSRPRGTSQSDISTLPTVPSSPDLVSETAMEVDTPAEQFLQSTSTSAQ 420
 QY 421 AHSTSSPESHSTPLSSPDSSEOROSVEAGHHTHOS----- 459
 DB 421 AHSTSSPESHSTPLSSPDSSEOROSVEAGHHTHOS----- 459
 QY 460 -----DNNN 463
 DB 460 -----DNNN 463
 QY 481 KAEQROQELAAHTQOQPTSDQSHEGSSQDPHASPSSVYVKNQKLSMLDEQODNN 540
 DB 481 KAEQROQELAAHTQOQPTSDQSHEGSSQDPHASPSSVYVKNQKLSMLDEQODNN 540
 QY 524 EKLSPKPCTGEPVLSIHTSTGTTSTKILNFTDEWSSIASRIGSHCKSEGOEESFV 523
 DB 524 EKLSPKPCTGEPVLSIHTSTGTTSTKILNFTDEWSSIASRIGSHCKSEGOEESFV 523
 QY 541 EKLSPKPCTGEPVLSIHTSTGTTSTKILNFTDEWSSIASRIGSHCKSEGOEESFV 600
 DB 541 EKLSPKPCTGEPVLSIHTSTGTTSTKILNFTDEWSSIASRIGSHCKSEGOEESFV 600
 QY 584 POSSVQPPGDSSETKAPESESESDVTKYQGVSAENPVNHNITQSDFTAKPLDLSNGE 583
 DB 584 POSSVQPPGDSSETKAPESESESDVTKYQGVSAENPVNHNITQSDFTAKPLDLSNGE 583
 QY 601 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENNTNPPQOTATGSAHEE 643
 DB 601 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENNTNPPQOTATGSAHEE 643
 QY 644 TSTRSALODTDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 703
 DB 644 TSTRSALODTDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 703
 QY 721 TSTRSALODTDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 780
 DB 721 TSTRSALODTDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 780
 QY 741 IRRPLVKKVYKGRHSRTMIKEANFWANFVWSDCGHIFIMDRHTAEHLMLLEADNHV 763
 DB 741 IRRPLVKKVYKGRHSRTMIKEANFWANFVWSDCGHIFIMDRHTAEHLMLLEADNHV 763
 QY 764 VNCLOPHPPPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNEMLLETRNTIYP 823
 DB 764 VNCLOPHPPPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNEMLLETRNTIYP 823
 QY 824 ASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 860
 DB 824 ASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 860
 QY 901 ASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 937
 DB 901 ASFMLRLASLNHIRADRLLEGDRSESGSGQENENEDEE 937
 RESULT 4
 ABB05707
 ID ABB05707 standard; Protein; 482 AA.
 AC ABB05707;
 DT 30-APR-2002 (first entry)
 DE
 XX
 Query Match 30-APR-2002 (first entry)
 Best local similarity 30-APR-2002 (first entry)
 Matches 30-APR-2002 (first entry)

DE Human signal transduction protein clone test_11c22.
 XX Human: foetal brain; foetal kidney; melanoma; testis; amygdala;
 KW gene therapy; chromosome 1q23.2-24.3.
 XX
 XX Hemo sapiens.
 XX MO200198454-A2.
 XX 27-DEC-2001.
 XX
 XX 25-APR-2001; 2001WO-1B02050.
 XX
 XX 25-APR-2000; 2000US-199380P.
 XX
 XX (GENU-) GERMAN HUMAN GENOME PROJECT.
 XX
 XX Wiemann S;
 XX
 XX WPI: 2002-055860/07.
 DR N-PSDB: ABA93744.
 XX
 XX Human cDNA sequences and clones derived from human fetal brain, fetal
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
 PT screening and therapy -
 XX
 XX Claim 1; Page 287; 611pp; English.
 XX
 XX The present invention describes assemblies and computer readable media
 CC comprising novel human cDNA sequences and clones derived from human
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
 CC present invention which encode the proteins given in ABB05662 to
 CC ABB05729. The human cDNA sequences and clones can be used in gene
 CC therapy. The clones may be used in a variety of applications, for
 CC example they may be used in profiling assays, for providing large arrays
 CC of human genetic material for implementing large-scale screening
 CC strategies and for treating diseases via gene therapy procedures.
 CC
 XX Sequence 482 AA;
 SO
 Query Match 52.8%; Score 2390; DB 23; Length 482;
 Best Local Similarity 95.9%; Pred. No. 1.5e-166;
 Matches 462; Conservative 0; Mismatches 0; Indels 20; Gaps 1;
 QY 399 MEVTPAEOFLQFSTSTMSQAOSTSPSTPSHTPLSSPDSEQRQSVASGHHTHQ 458
 DB 1 MEVTPAEOFLQFSTSTMSQAOSTSPSTPSHTPLSSPDSEQRQSVASGHHTHQ 60
 QY 459 S-----DNNNEKLSPPKPGGEPLSLHYSTEGTTSTIKINFTDE 498
 DB 61 SDSPSSVYNKQSGMSLDEQDDNNNEKLSPPKPGGEPLSLHYSTEGTTSTIKINFTDE 120
 QY 499 WSSIASSRIGISHCKEAGESEFVPOSSVOPPGGDEETKAPESSEEDVTYQEGVSAEN 558
 DB 121 WSSIASSRIGISHCKEAGESEFVPOSSVOPPGGDEETKAPESSEEDVTYQEGVSAEN 180
 QY 559 PVEHHITTSQDKTARPLDSNNGERDNLNDRSCGVPESASSEKAKEPTSDQSTES 618
 DB 181 PVEHHITTSQDKTARPLDSNNGERDNLNDRSCGVPESASSEKAKEPTSDQSTES 240
 QY 619 ATNNNNNPPEPQTEATGSAHEETSRDASALODTDDSDDDPVLIGARYRAGPGRRS 678
 DB 241 ATNNNNNPPEPQTEATGSAHEETSRDASALODTDDSDDDPVLIGARYRAGPGRRS 300
 QY 679 AVARIOEFFRRRERKEMEELDTLNIIRPLVKMYKGRHSRTMIKEANFGANFVAGS 738
 DB 301 AVARIOEFFRRRERKEMEELDTLNIIRPLVKMYKGRHSRTMIKEANFGANFVAGS 360
 QY 739 DCGHIFIMDHTAEHMLLLEADNNVYVNCLOPHRDPPLIASSGIDYDIKINSPLEESRTFN 798
 DB 361 DCGHIFIMDHTAEHMLLLEADNNVYVNCLOPHRDPPLIASSGIDYDIKINSPLEESRTFN 420

QY 799 RKLADENVITRNLMEETRTITVPASFMLRMLASINHLRADRLGDRSGSCQENED 858
 DB 421 RKLADENVITRNLMEETRTITVPASFMLRMLASINHLRADRLGDRSGSCQENED 480
 QY 859 EE 860
 DB 481 EE 482
 RESULT 5
 AAU19599
 ID AAU19599 standard; Protein; 301 AA.
 XX
 XX AAU19599;
 AC
 XX
 XX 06-DEC-2001 (first entry)
 DE
 XX Human diagnostic and therapeutic polypeptide (DTRP) #185.
 XX
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 OS
 XX Hemo sapiens.
 XX
 XX W0200162927-A2.
 XX
 XX 30-AUG-2001.
 XX
 XX 21-FEB-2001; 2001WO-US06059.
 XX
 XX 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 12-MAY-2000; 2000US-0185216.
 PR 15-MAY-2000; 2000US-0203785.
 PR 16-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PI Panzer SF, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'sa SA, Amhey S, Dahl CR, Daniels SE;
 PI Difour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SK, Chen W;
 PI Cohen HU, Hodgson DM, Lincoln SE, Jackson SJ;

XX WPI: 2001-502867/55.
DR N-PSDB: AAS31170.

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics

PS Claim 27; Page 507; 522pp; English.

XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DTRP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DTRP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DTRPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DTRPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DTRPs and in assays to identify modulators of DTRP
CC expression and activity. The anti-DTRP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DTRP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DTRPs in samples (e.g. by enzyme linked immunosorbent
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
CC therapeutic (DTRP) polypeptides of the invention.

XX Sequence 301 AA:

Query Match 32.2%; Score 1459; DB 22; Length 301;
Best Local Similarity 95.0%; Pred. No. 1.3e-98;
Matches 283; Conservative 0; Mismatches 1; Indels 14; Gaps 1;

QY 459 SDNNKSLSPKRGTEPVLSTHYSTEGTTSTIKNTFDWSSIASSSSGTSHCKSEQ 518
DB 4 SDNNKSLSPKRGTEPVLSTHYSTEGTTSTIKNTFDWSSIASSSSGTSHCKSEQ 63
QY 519 EESFVPOSSVOPPEGDSSTKAPESSEEDVTCKOEGVSAENPVENHINITOSDKFTAKPLD 578
DB 64 EESFVPOSSVOPPEGDSSTKAPESSEEDVTCKOEGVSAENPVENHINITOSDKFTAKPLD 123
QY 579 SNSGRNDLNDRCGVPESASSEKAKEPETSDDTSTESATNENNTNPEFOFOTBACGP 638
DB 124 SNSGRNDLNDRCGVPESASSEKAKEPETSDDTSTESATNENNTNPEFOFOTBACGP 183
QY 639 SAHEETSPRDSALDPTDSDDDPVLIPGARVRAGPGD-----RRAVARIO 684
DB 184 SAHEETSPRDSALDPTDSDDDPVLIPGARVRAGPGDPRNINIGTTIGDIRMRSARVARIQ 243
QY 685 EEFRRRRKKEKEEELDTNIRRLPVKMYKGRNSRTMIKEANFGANFVMSGSDCGH 742
DB 244 EEFRRRRKKEKEEELDTNIRRLPVKMYKGRNSRTMIKEANFGANFVMSGSDCGH 301

RESULT 6
AAB92537
ID AAB92537 standard; Protein: 163 AA.

AC AAB92537;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:10703.

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUN-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-7) HELIX-RBS INST.

XX Ota T, Isogai S, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8; SEQ ID 10703; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 163 AA:

Query Match 18.7%; Score 847; DB 22; Length 163;
Best Local Similarity 98.2%; Pred. No. 3.6e-54;
Matches 162; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 696 MEELDTNIRRLPVKMYKGRNSRTMIKEANFGANFVMSGSDCGHIFIMRHTAEHLM 755
DB 1 MEELDTNIRRLPVKMYKGRNSRTMIKEANFGANFVMSGSDCGHIFIMRHTAEHLM 60

QY 756 LLEADNHVNCIOPHPFDITLASSGIDYDIKTIWSPLEESRIRNRKLADEVITRNELMLEE 815
DB 61 LLEADNHVNCIOPHPFDITLASSGIDYDIKTIWSPLEESRIRNRKLADEVITRNELMLEE 118

QY 816 TRNTITVPASFMLRLMLASLNHTRADRLGDRSGSGQENENDEDE 860
DB 119 TRNTITVPASFMLRLMLASLNHTRADRLGDRSGSGQENENDEDE 163

RESULT 7

AAB27984
ID AAB27984 standard; Protein; 129 AA.
XX
AC AAB27984;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein BLAST search protein SEQ ID NO: 138.
XX
KM Cytostatic; immunosuppressive; neutropenic; neuroprotective; antiviral;
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KM vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PY WO200055171-A1.
XX
HD 21-SEP-2000.
XX
PE 09-MAR-2000; 2000WO-US06043.
XX
PR 12-MAR-1999; 99US-0124146.
XX 23-NOV-1999; 99US-0167061.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638174/61.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; Page 400; 438pp; English.
XX
CC The invention relates to the isolation of genes AAC59049-C59098 encoding
CC the human secreted proteins AAB27907-B27956. This sequence represents a
CC fragment of the protein encoded by the gene given in the descriptor
CC line. The sequence is used as a query sequence for doing BLASTX searches
CC to determine homologous sequence to the protein. The genes and proteins
CC are useful for preventing, ameliorating or treating medical conditions,
CC e.g. by protein or gene therapy. The genes are isolated from a range of
CC human tissues disclosed in the specification. The nucleic acids,
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,
CC and other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple
CC sclerosis, rheumatoid arthritis and ulcerative colitis; (c) wound
CC cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and
CC (f) infectious diseases such as viral, bacterial, fungal and parasitic
CC infections.
XX
SQ Sequence 129 AA;
XX
Query Match 15.2%; Score 688; DB 21; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.1e-42;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 SLNHIRADR 129
RESULT 8
ID AAB27926
XX
AC AAB27926 standard; Protein; 127 AA.
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein SEQ ID NO: 80.
XX
KM Cytostatic; immunosuppressive; neutropenic; neuroprotective; antiviral;
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KM vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PY WO200055171-A1.
XX
HD 21-SEP-2000.
XX
PE 09-MAR-2000; 2000WO-US06043.
XX
PR 12-MAR-1999; 99US-0124146.
XX 23-NOV-1999; 99US-0167061.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-638174/61.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 11; Page 371; 438pp; English.
XX
CC Sequences AAB27907-B27956 represent the amino acid sequences of 50
CC human secreted proteins encoded by the genes AAC59049-C59098. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 127 AA;
XX
Query Match 14.7%; Score 664; DB 21; Length 127;
Best Local Similarity 100.0%; Pred. No. 6.4e-41;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 ENEDEE 860
 DB 121 ENEDEE 126

RESULT 9

ID ABB29956 standard; Peptide: 110 AA.

AC ABB29956;

DT 01-FEB-2002 (first entry)

DE Peptide #2607 encoded by breast cell single exon nucleic acid probe.

KW Human; microarray; single exon probe; gene expression; breast; disease; cancer.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 12924; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labeled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 110 AA;

XX Query Match 12.6%; Score 569; DB 22; Length 110;

XX Best Local Similarity 100.0%; Pred. No. 4.7e-34;

XX Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGSRNDLNDRCSCGVPESASSEKAKEPETSQSTESATNENNTNP 627

DB 1 QSDKFTAKPLDLSNGSRNDLNDRCSCGVPESASSEKAKEPETSQSTESATNENNTNP 60

QY 628 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 676

DB 61 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 109

RESULT 10

ID ABB20555 standard; Protein: 110 AA.

AC ABB20555;

DT 23-JAN-2002 (first entry)

DE Protein #2554 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system; cardiovascular disease; hypertension; cardiac arrhythmia; congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human hearts

XX Claim 15; SEQ ID NO 22325; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see AB21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays.

XX By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 110 AA;

XX Query Match 12.6%; Score 569; DB 22; Length 110;

XX Best Local Similarity 100.0%; Pred. No. 4.7e-34;

XX Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGSRNDLNDRCSCGVPESASSEKAKEPETSQSTESATNENNTNP 627

DB 1 QSDKFTAKPLDLSNGSRNDLNDRCSCGVPESASSEKAKEPETSQSTESATNENNTNP 60

QY 628 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 676

DB 61 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 109

RESULT 11

AAM55960
ID AAM55960 standard; Protein; 110 AA.

AC AAM55960;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28065.

KW Human; brain expressed exon; gene expression analysis; probe;
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-483446/52.

PS Single exon nucleic acid probes for analyzing gene expression in human

PS brains -

PS Example 4; SEQ ID NO: 28065; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 110 AA;

XX Query Match 12.6%; Score 569; DB 22; Length 110;

XX Best Local Similarity 100.0%; Pred. No. 4.7e-34;

XX Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28634.

KW Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukemia; lymphoma; myeloma.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-488900/53.

PS Human genome-derived single exon nucleic acid probes useful for

PS analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 28634; 658bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukemia and myeloma. The present sequence is a

XX protein encoded by one of the probes of the invention.

XX Sequence 110 AA;

XX Query Match 12.6%; Score 569; DB 22; Length 110;

XX Best Local Similarity 100.0%; Pred. No. 4.7e-34;

XX Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX RESULT 13

XX AAM16144

XX ID AAM16144 standard; Protein; 110 AA.

XX AC AAM16144;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #2578 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX KW cervical cancer.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00670.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -
 PS Claim 27; SEQ ID NO 20970; 487pp; English.
 CC The present invention relates to human single exon nucleic acid probes
 CC (SNP: see A110068-A1128459). The present sequence is a peptide encoded
 CC by one such probe. The SNPs are derived from human Hela cells. The SNPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 110 AA;
 Query Match 12.6%; Score 569; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 4.7e-34;
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 568 OSDKFTAKPLDLSNGSRDNLNDRSCGVPEESASSEKAKEPTSDQTSSTESATNENNTNP 627
 DB 1 OSDKFTAKPLDLSNGSRDNLNDRSCGVPEESASSEKAKEPTSDQTSSTESATNENNTNP 60
 QY 628 EPQFQTEATGSAHEETSTRDSALODTDDSDDDPVILPGARYRAGGDR 676
 DB 61 EPQFQTEATGSAHEETSTRDSALODTDDSDDDPVILPGARYRAGGDR 109
 RESULT 14
 AAM03874
 ID AAM03874 standard; Protein; 110 AA.
 AC AAM03874;
 XX
 DT 09-OCT-2001 (first entry)
 DE Peptide #556 encoded by probe for measuring breast gene expression.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 XX in a human breast -
 PS Claim 27; SEQ ID NO 12614; 322pp; English.
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see A110010-A110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer; disorders of development,
 CC inflammatory diseases of the breast; fibrocystic changes; proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 110 AA;
 Query Match 12.6%; Score 569; DB 22; Length 110;
 Best Local Similarity 100.0%; Pred. No. 4.7e-34;
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 568 OSDKFTAKPLDLSNGSRDNLNDRSCGVPEESASSEKAKEPTSDQTSSTESATNENNTNP 627
 DB 1 OSDKFTAKPLDLSNGSRDNLNDRSCGVPEESASSEKAKEPTSDQTSSTESATNENNTNP 60
 QY 628 EPQFQTEATGSAHEETSTRDSALODTDDSDDDPVILPGARYRAGGDR 676
 DB 61 EPQFQTEATGSAHEETSTRDSALODTDDSDDDPVILPGARYRAGGDR 109
 RESULT 15
 ABG37880
 ID ABG37880 standard; Peptide; 110 AA.
 AC ABG37880;
 XX
 DT 19-AUG-2002 (first entry)
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 27545.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.

30-JAN-2001; 2001WO-US00665.
 04-FEB-2000; 2000US-180312P.
 26-MAY-2000; 2000US-207456P.
 30-JUN-2000; 2000US-0608408.
 03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-234687P.
 27-SEP-2000; 2000US-236359P.
 04-OCT-2000; 2000GB-0024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2002-114183/15.
 Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -
 Claim 27; SEQ ID NO 27545; 634bp; English.
 The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridize at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagazer syndrome, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 110 AA;
 Query Match 12.6%; Score 569; DB 23; Length 110;
 Best Local Similarity 100.0%; Pred. No. 4,7e-34;
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

568 QSDKFTAKPLDSNGERDNLNDRSCGVPESASSEKAKEPETSDDTSTESATNENNTNP 637
 DB 1 QSDFTAPLPLDSNGERDNLNDRSCGVPESASSEKAKEPETSDDTSTESATNENNTNP 60
 628 EPQGTETATGPSAHEETSTRDSDALQDTDDSDDDPVLIPGARFRAGGDR 676

DB 61 EPQGTETATGPSAHEETSTRDSDALQDTDDSDDDPVLIPGARFRAGGDR 109
 RESULT 16
 ABB61280
 ID ABB61280 standard; Protein; 628 AA.
 AC ABB61280;
 DE 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 10632.
 DE Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li FWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL05383.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
 PT
 PS Disclosure; SEQ ID NO 10632; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB101840-AB130511).
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 628 AA;
 Query Match 11.2%; Score 508.5; DB 22; Length 628;
 Best Local Similarity 23.2%; Pred. No. 1.8e-28;
 Matches 191; Conservative 107; Mismatches 261; Indels 265; Gaps 26;

10 LHMVRRKRSGLDEPSRLRSRYIGRRERIORLKLATLVNHDGCVNTICMNDTGEYIISG 69
 DB 28 LHM--ORQYGHLEDEMLRRRLASPAVYDRLEDAVIVGHEGVCNLEMTDGMWLASG 85
 70 SDDTKLVIISNPSRKVLTITRSGRANIFSAKFLPCTNDKQIVSGGQVIRYTVNEODA 129
 DB 86 SDDRVVIMIPDFRKRLVAVITKHLGNVFSYKFLPKTNNISVATCAAKRFIYVINDPN 145
 130 ETNRQCOFTCHGYTTEIMTVPNNDPYFLSCGEGDGVWFPDTRIKTSCTEKDCCKDILIN 189
 DB 146 ETTFESC--ICHFSSRAKRLATQDSPHYWVSAGEDCIQLDIRPHRCRPEEGICVRLN 203
 190 C-----RRAATSVACIPPIPIYLAIVGCDSSVRIYDRMLGRATGAGTGTGMARF 244

```

Db 204 LHDQLENTAKCLAIINPRFTEYLAVGTNDPFARVYDRRLPST-----NGNGISACVAYY 258
OY 245 IPSHLNKKSCR-----VTSICSEDOGELIVSYSDIYLFDPKDDTARELKTPSAE 296
Db 259 AAGQIVKNISRNIVHEPRITLITLFGNGTELLVNICGEHYRFDLNH-----306
OY 297 ERREELROPVVKRLRLGDMWSDTGPRAPRESEBERDEGSPNVSLMQRMSDMLSRWFEFA 356
Db 307 -----AEPV-----FYDLPATSTLVHEE 326
OY 357 SEVAQSNRGRGRSRPGTSGSDISTLPVPSPDLEVESETAMEVDTPAEQLOPSTST 416
Db 327 EPVKMPHRSR-----SLPT-----ELEVK-----KEGNDFFE-----354
OY 417 MSAQAHSTSTSPSTPLSSPDSEOROSVSAHGHTHOSDNNNEKLSRPGTGPEV 476
Db 355 -----NGKLVDAIDAVSALAKYPOE-----Y 377
OY 477 LSLHSTEGTWTSTKLN-FTDEWSSIASSSNGISGSHCKSEGOESFVPOSSVQPEGDS 535
Db 378 LVLN-----RATALMRGWFEGDIYALRDCHEAL-----RLDP-----410
OY 536 ERKAPESSEDTKYQEGVSAENPVENHINIQSDKFTAKPLDSNGERNDLNDSGCV 595
Db 411 -----SYVAHRLA-----RALLELR-----428
OY 596 PESASSEKAKEPETSQDSTESATENNNTNPPQFOTATGSAHEETSTRDALSQDPTD 655
Db 429 PQDADCCLOALTOREFDFANNHGVMLNDIKENRQSKS--PEAE-----LQPE 478
OY 656 DSDDDPVLIPGARYRAGPDRSAVARIOEFFRRKKERKEMELDLNLNRLPVKKVYKG 715
Db 479 VDD-----GFRYL-----RKMEYDLSRSTARDYMQR-YYG 508
OY 716 HNSRFTIMEKANPWA--NFVMSGSDCGHLEFMDRTAHEIMLEADNHNVCLOHPFD 773
Db 509 KCVNTTIDIEANLISGGEFIAAGSDGKMYTIEGDTGKRAYVRADSAIVNCVQHPST 568
OY 774 PTLASSGIDYDIKIVSPLEESRIFNKLADDEVITRNELMEETR 817
Db 569 CMLATSGIDHNKIKMSPCAAS-----AE-----RNLVADVTR 602

RESULT 17
AAU80462
ID AAU80462 standard; Protein; 628 AA.
XX
AC AAU80462;
XX
DT 12-MAR-2002 (first entry)
XX
DE Fruit fly adipose protein, adp, #2.
XX
KW Adipose protein; adp; obesity; transgenic animal; obesity;
KW adipositas; bulimia; wasting; cachexia; eating disorder;
KW body weight disorder; weight loss; cancer; infectious disease;
KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;
KW hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;
KW gastrointestinal diseases; inflammatory bowel disease;
KW ulcerative colitis; anorexia nervosa; glycogen storage disease;
KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;
KW infertility; acquired immunodeficiency syndrome; AIDS.
XX
XX Drosophila melanogaster.
XX
PN WO200196371-A2.
XX
PD 20-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-EP06713.
XX
PR 16-JUN-2000; 2000US-211914P.
PR 23-JUN-2000; 2000EP-0113049.

```

```

PR 28-JUN-2000; 2000US-214518P.
PR 17-APR-2001; 2001EP-0109537.
XX
PA (DEVE-) DEVELOGEN AG.
XX
PI Broenner G, Ciossek T, Dohrmann C, Haeder T, Rothe M;
XX WPI; 2002-106464/14.
XX N-PSDB; ABR16345.
XX
PT Novel nucleic acid encoding adipose polypeptide which regulates, causes
PT or contributes to obesity, useful for treating obesity, heart disease,
PT hypertension, infertility, and controlling weight loss in cancer
PT patients.
XX
PS Claim 1; Page 142-144; 188pp; English.
XX
CC The invention relates to a nucleic acid encoding a adipose (ADP)
CC polypeptide which regulates, causes or contributes to obesity in an
CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,
CC modulators of adp activity, adp antisense nucleic acids, expression
CC vectors, adp transgenic animals are useful in the diagnosis and
CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating
CC disorders and/or disorders of body weight/body mass, weight loss due to
CC cancer or infectious diseases, genetic disorders associated with
CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,
CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,
CC gastrointestinal diseases, inflammatory bowel disease, ulcerative
CC colitis, and anorexia nervosa. They are also useful for treating
CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid
CC storage diseases and for treating lipomas, and/or liposarcomas. The
CC compositions are also useful for treating heart disease, hypertension,
CC and infertility and for treating conditions associated with under weight
CC e.g. enhancing or controlling fertility, controlling weight loss in
CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The
CC present sequence is an adp protein.
XX
SQ Sequence 628 AA;
XX
Query Match 11.2%; Score 508.5; DB 23; Length 628;
Best local Similarity 23.2%; Pred. No. 1.8e-28;
Matches 191; Conservative 107; Mismatches 261; Indels 265; Gaps 26;
OY 10 LMDVRRKSLGLEDPRLNSRYLGRREFIQRLKLEATLVNHDGCVNTICWNTGEXYLSG 69
Db 28 LHM--ORQGHLEDEMLRRRLASPAVYDRLBOEAVLVGHECVCLLEWTGGMMLASG 85
OY 70 SDDTKIVISNPIYRKVLTITRSGRANIPSAKFLPCTNDKQIYVSGSDGYIFFTNVEQDA 129
Db 86 SDDYRWIMDPFRKKLVHIVIRTHLGNVESGKFLPRTNNSIVATCAADKRTIYYDINDPN 145
OY 130 ETNRQCFCHGYTTEIMTVPNDPYFLSGGEGTVRWEDTRIKTSCTEKDKDILLN 189
Db 146 ETFLSC--ICHFSRAKRIANAOQSPHFWAGSDGCLQDIDIEPRRCREBIBGVRLN 203
OY 190 C-----RRAATVSLCOPPIPYLAVGSSSVIRIDRMILSTRATGNYAGRTTGVARF 244
Db 204 LHDQLENTAKCLAIINPRFTEYLAVGTNDPFARVYDRRLPST-----NGNGISACVAYY 258
OY 245 IPSHLNKKSCR-----VTSICSEDOGELIVSYSDIYLFDPKDDTARELKTPSAE 296
Db 259 AAGQIVKNISRNIVHEPRITLITLFGNGTELLVNICGEHYRFDLNH-----306
OY 297 ERREELROPVVKRLRLGDMWSDTGPRAPRESEBERDEGSPNVSLMQRMSDMLSRWFEFA 356
Db 307 -----AEPV-----FYDLPATSTLVHEE 326
OY 357 SEVAQSNRGRGRSRPGTSGSDISTLPVPSPDLEVESETAMEVDTPAEQLOPSTST 416
Db 327 EPVKMPHRSR-----SLPT-----ELEVK-----KEGNDFFE-----354
OY 417 MSAQAHSTSTSPSTPLSSPDSEOROSVSAHGHTHOSDNNNEKLSRPGTGPEV 476

```

Db 355 -----NGKLVDALDAYSAALAKYPOGE-----Y 377

Qy 477 LSLHSTECTTTSTIKLN-FTDEWSSIASSSGIGSHCKSEGOEESFVQSVQPEEGDS 535

Db 378 LYL-----RATLMRRGFGDIYALRDCHEAL-----RLDP-----410

Qy 536 ETKAPEESEDVTKYQEGSAENPVENHININOSDKFTAKPLDLSNGERNDLMDSCGV 595

Db 411 -----SYKAHFRLLA-----RALLELR-----428

Qy 596 PESASSEKAKEPETSDQSTESATENNENTNEPOFOTATGSAHEETSTRDALODTD 655

Db 429 PQDAPQOCALQIRPFDPFANNHGVLMNDIKENRQSS--PEAE-----LQPYE 478

Qy 656 DSDDDPVILPGARYAGPDRSAAVARIOEPPRRKREMEELDTLIRPLVKVYKG 715

Db 479 VDD-----GFRYL-----RLKMEYDTRSTARDWQR--YVG 508

Qy 716 HRNRTMIKEANFWGA--NFWSSGDCGHIFIMDRHTAHMLLEADNHVYVNCLOPPPD 773

Db 509 HCNVTTDIKEANYLSGGEFTIAGSDGNNYIMEGDTGIRAVYRADSAIVNCVQPHPI 568

Qy 774 PILASSGIDYDIKIMSPLEESRIFNRKLADEVITRNELMLEETR 817

Db 569 CMLATSGIDHNKIMKPCAS-----AE-----RNLVADYTR 602

RESULT 18

AAU80464

ID AAU80464 standard; Protein; 677 AA.

XX

AC AAU80464;

DT 12-MAR-2002 (first entry)

XX

DE Human adipose protein, adp.

XX

KW Adipose protein; adp; obesity; transgenic animal; obesity; adipositas; bulimia; wasting; cachexia; eating disorder; body weight disorder; weight loss; cancer; infectious disease; hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome; hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder; gastrointestinal diseases; inflammatory bowel disease; ulcerative colitis; anorexia nervosa; glycogen storage disease; lipid storage disease; lipoma; liposarcoma; heart disease; hypertension; infertility; acquired immunodeficiency syndrome; AIDS.

KM

XX

OS Homo sapiens.

XX

PN WO200196371-A2.

XX

PD 20-DEC-2001.

XX

PF 13-JUN-2001; 2001WO-EP06713.

XX

PR 16-JUN-2000; 2000US-21914P.

PR 23-JUN-2000; 2000EP-0113049.

PR 28-JUN-2000; 2000US-214518P.

PR 17-APR-2001; 2001EP-0109537.

XX

PA (DEVE-) DEVELOPEN AG.

XX

PI Broenner G, Ciossek T, Dohrmann C, Haeder T, Rothe M;

DR WPI; 2002-106464/14.

DR N-PSDB; ABK16347.

XX

PT Novel nucleic acid encoding adipose polypeptide which regulates, causes or contributes to obesity, useful for treating obesity, heart disease, hypertension, infertility, and controlling weight loss in cancer patients -

PT

XX

PS Claim 1; Page 153-155; 188pp; English.

XX

CC The invention relates to a nucleic acid encoding a adipose (ADP)

CC polypeptide which regulates, causes or contributes to obesity in an

CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,

CC modulators of adp activity, adp antisense nucleic acids, expression

CC vectors, adp transgenic animals are useful in the diagnosis and

CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating

CC disorders and/or disorders of body weight/body mass, weight loss due to

CC cancer or infectious diseases, genetic disorders associated with

CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,

CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,

CC gastrointestinal diseases, inflammatory bowel disease, ulcerative

CC colitis, and anorexia nervosa. They are also useful for treating

CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid

CC storage diseases and for treating lipomas, and/or liposarcomas. The

CC compositions are also useful for treating heart disease, hypertension,

CC and infertility and for treating conditions associated with under weight

CC e.g. enhancing or controlling fertility, controlling weight loss in

CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The

CC present sequence is an adp protein.

XX

SQ Sequence 677 AA;

Query Match 11.08; Score 499.5; DB 23; Length 677;

Best Local Similarity 23.68; Pred. No. 9.4e-28;

Matches 209; Conservative 116; Mismatches 297; Indels 265; Gaps 33;

Qy 13 DVRRSLGLEDPRLSRSLRGREFIORKLEATLVNHDGCVNTICMNDTGYEYISGSD 72

Db 9 DLIRQIKERALSFRERYHTDPTIRKLGAELOHSGCVNCEMNEKGDILASGSD 68

Qy 73 TKLVISNPEYSKRVLTIRSGHRANIFSAKFLPTNDKROIVSCSGGVLYETVVEDQAEIN 132

Db 69 QHITWDEPLHKKILSMHTGTANIFSVKFLPHADRLITGAADSKVHVDLYKETIH 128

Qy 133 RQCFCHYGTTELMYVNDPYTFLSCGEQSTVWFPTIRITSCTECKC--DDILIN-- 189

Db 129 --MGDTNRVYKRIATPMPMPTFWSAEDLILQYDILR-----ENSKASEVILDTL 178

Qy 190 --CRR--AATSVATCPPIYTLAVGSDSSVRYDRLM-----GTRATGNYVGRG-- 237

Db 179 EYCGQIVAKCLTVNPQONNCLAVASGPYVLYIRIMHNHNRKMSKQSPSAGVHTFCR 238

Qy 238 -----TGMAVAFIDSHL-----NNKSCRV--TSLCYSEDOEILVSYSDIYLFDPK 283

Db 239 QKPLPDGAQYVYVAGHLPVKLPDYNNRLVATVYTFSPNCTELLVNNGGQVYLF-- 296

Qy 284 DDTARELKTPEAERREELRQPPVYKRLRLRGWSDTGPAPARESERERGBQSPNVSLMQ 343

Db 297 -----LTYKORPYTFLL-----PKCHSGSEVONGKMNSTN----- 326

Qy 344 RMSDMLSRPEEASEVVASNNGRGRSRPRGTSQSDISTLPVPSPDLE--VSETAMEYD 402

Db 327 GVSNGVSGMLHNSGFLPESRGHVSPQ-----VELPPLYERKQOQANEA-- 372

Qy 403 TPAQOFLQPTSSYMSAQAHSSTSPESHPTPLISSPDSQROSVASGHTTHQSDNN 462

Db 373 FACQOW-----TQALDLYS--KAVQAPAHNMALYG-----NR 402

Qy 463 NEXLSRPGCEPVLISLHYSBEGTTSTIKNFTDEWSSIASSSGIGSHCKSEGOEESF 522

Db 403 AAAYMKRMQSD-----HYDALRDLKALISLN-----PCHLKA----- 435

Qy 523 VPOSSVQPEEGSETKAPEESEDVTKYQEGSAENPVENHININQ-----SDKFTAKPLDS 579

Db 436 -----HRLAKCLFELKYVAEAL-- 453

Qy 580 NSGERNDLNDKSCGVPEASASSEKAKEPETSDQSTESATENNENTNEPOFOTATGSPS 639

Db 454 -----ECIDDFKGRPEQAHSSA-----CDALG-- 476

Qy 640 AHETSTRD--SALQDIDDSDDDPVLLPGAVYRAGPGRRAVAVARIOEPPRRKREKME 697

Query Match	Best Local Similarity	Matches 193;	Conservative	Score 470.5;	DB 23;	Length 580;
10.48;	24.38;	95;	Mismatches 251;	Pred. No. 9.9e-26;	Indels 255;	Gaps 31
40	RLKLEATLVNHDGVNFTICWNDTGEVILSSDDTQTKLVISNRYKVLTTIRSGHRNIFTS	99				
1	RLGLEALDQSHSGCVNLLENNEKGDLLASSDDQNTIWMVRLNHNKLLSMNHTGANIFTS	60				
100	AKFLPTCTNDKQIVSCSGDGYIFFTNVQDDETRNOCQFTGHNIGTYYTILVMPNDPTFLS	159				
61	VKFLPHAGNDLITLGCADSKSVNHNDLVKFTIH---	117				
160	CGEGCTVWRFEDTRTKTSCTEKDC--DDILN---	112				
118	AAEDGLRKQYDLR-----ENSKHSEVLLDILEYCGQLYEAKCLTVNPPDNNKLANGAS	170				
213	DSRYIVTIDRRLI-----GTRATGNVAGRGY-----	253				
171	GFPRVLRVDIRINHNHRSKMSKQSPAGVNTFCDRKQPLRDGAAQYVAGVGHPLVVKLDYNNR	230				
254	CRV---TSLCYSEGGQELIVSYSSDIYIDRPKQDTARELKTFRPAERRELRQPPVKRL	310				

```

Db      231 LRLVATYVFSPRGTELLVLMVGEGVYTFD-----LTYKRPYTF 272
QY      311 RLRLDMEDTQGRAPREERERDEGOSPNYSLMQMSDMLSRFPEEASVQASNRGRS 370
Db      273 L-----PRCHSSGEVQNGKMSIN-----GVNGYSNGLIHLSNFRLEPESRGIVS 318
QY      371 PRGETSODISITLTPVSPDLE-VSETAMEVDTPAEQFLQPSTSTMSAQHSTSSPTE 429
Db      319 PQ-----VELPYLERVKQQAWEA-FACQOW-----TQAIDLYS-KAVQR 356
QY      430 SPHSPTLLSSPDSQORSVEASGHHTHHQSDDNNKELSPKPGTGEFVLSLHYTEGTTTS 489
Db      357 APHNAMLTG-----NRAAMTKRMKD-----HYDALRDLCK 389
QY      490 TIKLFTDMSGLASSRSKIGSHCKSBQESFVPOSSVQPRGDSETAPESSSDVTK 549
Db      390 AILSL-----PCHLKA-----400
QY      550 YQEGVSAENPEVNINITQ---SPKFAKPLDSSNGERNDLDRSCGVPESSASEKAK 606
Db      401 -----HRLARCLFELKYAEAL-----BOLDPFKK 427
QY      607 EPEISDQSTESATNENTNPEQFQTEATGAPSAHEETSTRD--SALQDTDSDDPVL 664
Db      428 PEGQAHSSA-----CDALG-----RDITALEFSKNDGE 456
QY      665 PGARYRAGCGDRRSVARIQIEFFRRKRKEMELDLINRLRP--VKMYKKHRRSRKM 722
Db      457 -----KKGGG--GAPVPL-----RSTRSKDISDEDEVLRENSYDQFRIQCHCNTTD 504
QY      723 IKEANFMGAN--FYVAGSDCGHIFIMDRHTAEHLMLLEADNHVYVNCIQHPFPIILASSG 780
Db      505 IKEANFPGSMAQIYVSGSDDSFFIWEKETNLNLRVLQGESYVNCIQHPHYSCFLATSG 564
QY      781 IDYIKIMSPLES 794
Db      565 IDPVVRLMNPRES 578

RESULT 22
AA080469
ID      AA080469 standard; Protein: 580 AA.
XX
AC      AA080469;
XX
DT      12-MAR-2002 (first entry)
XX
DE      Mouse adipose protein associated protein.
XX
KW      Adipose protein; adp; obesity; transgenic animal; obesity;
KW      adipositas; bulimia; wasting; cachexia; eating disorder;
KW      body weight disorder; weight loss; cancer; infectious disease;
KW      hypoadonism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;
KW      hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;
KW      gastrointestinal diseases; inflammatory bowel disease;
KW      ulcerative colitis; anorexia nervosa; glycogen storage disease;
KW      lipid storage disease; lipoma; liposarcoma; heart disease; hypertension
KW      infertility; acquired immunodeficiency syndrome; AIDS.
XX
OS      Mus musculus.
XX
PN      W0200196371-A2.
XX
PD      20-DEC-2001.
XX
PE      13-JUN-2001; 2001WO-EP06713.
XX
PR      16-JUN-2000; 2000US-211914P.
PR      23-JUN-2000; 2000EP-0113049.
PR      28-JUN-2000; 2000US-214518P.
PR      17-APR-2001; 2001EP-0109537.
XX

```

PA (DEVE-) DEVELOGEN AG.
 XX
 PI Broenner G, Ciossek T, Dohmann C, Haeder T, Rothe M;
 XX WPI; 2002-106464/14.
 DR
 XX Novel nucleic acid encoding adipose polypeptide which regulates, causes
 PT or contributes to obesity, useful for treating obesity, heart disease,
 PT hypertension, infertility, and controlling weight loss in cancer
 PT patients -
 PS
 XX Disclosure; Page 165-167; 188pp; English.
 CC The invention relates to a nucleic acid encoding a adipose (ADP)
 CC polypeptide which regulates, causes or contributes to obesity in an
 CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,
 CC modulators of adp activity, adp antisense nucleic acids, expression
 CC vectors, adp transgenic animals are useful in the diagnosis and
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating
 CC disorders and/or disorders of body weight/body mass, weight loss due to
 CC cancer or infectious diseases, genetic disorders associated with
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative
 CC colitis, and anorexia nervosa. They are also useful for treating
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid
 CC storage diseases and for treating lipomas, and/or liposarcomas. The
 CC compositions are also useful for treating heart disease, hypertension,
 CC and infertility and for treating conditions associated with under weight
 CC e.g. enhancing or controlling fertility, controlling weight loss in
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The
 CC present sequence is an adp associated protein included in the
 CC sequence listing but not mentioned any where else in the specification.
 XX
 SQ Sequence 580 AA:
 Query Match 10.2%; Score 460.5; DB 23; Length 580;
 Best Local Similarity 24.1%; Pred. No. 5.4e-25;
 Matches 192; Conservative 92; Mismatches 252; Indels 261; Gaps 30;
 QY 40 RLKLEATLVHDCGVNTICWMDTGEYILSGSDPTFKLVISNPSRKVLITIRSGHRANIPS 99
 DB 1 RLGLFALOGHSGCVNCLNEMKEKDLASGSDQHTIYWDPLHKKLSMHTGHRANIPS 60
 QY 100 AKFLPCTNDQIVSCSGDGVLYITNVEDAETNRQCOFTCHYGYEIMTYPNDEPYTFLS 159
 DB 61 VKFLPHAGDRILITGADSKVHVDLTVYKETIH--MFGDHTNRVKRLATAPMWTNTEMS 117
 QY 160 CGEDGYRWFDTRIKTSCTKEDCK-DLLINCR-----AATSAICPPIRYIAVGCS 212
 DB 118 AAEEDGLROYDLR-----ENSKHSEVLIDLTEYCGPMEVAKCLTVNPQDNCLAVGAS 170
 QY 213 DSSVRIYIDRRML-----GTRATGNVAGRGT-----TGNVARIPIPSHL-----NNKS 253
 DB 171 GPVRYLIDIMINHKRSMQSPSAGVHTFCDROKRLPDGAQYVAGHLPPKLPDYNSR 230
 QY 254 CRV---TSLCYSEDEGILVSYSDIYLPKDKDTARELKTTPSAERREELROPPVKRL 310
 DB 231 LRVLYATYTFESPNGTELVVMGGEQYVLLD-----LTYKQRPYPTL 272
 QY 311 RLKGDMSDTGPRAPESEEREDGEOSPVSLMQRASDMLSRWFEEASEVYAOSSNRGRNSR 370
 DB 273 L-----PR-KCHSEVGVNGKMKSTN-----GVSNGVSNGLHLHS-----NGFRL 309
 QY 371 PRGTSOSDISTLPTVSSPDL-VSETAMEVDPFAEFLDPSTSTSAOAHSSSPTE 429
 DB 310 PE---SKGCS--PQVELPYLERKQOANA-FAQQOM-----TAIQLYS--QAVOK 355
 QY 430 SPHSTPLLSPEDEQROSVASGHTHQSNDNNKELSPKGTGEPVLSLHSTGTTTS 489
 DB 356 APHNAMLY-----NRAAAYMKRKMWDG-----HYDALRDCIK 388
 QY 490 TIKLNTDEMSSIASSSRGICSHCKSEGOESFVQSSVQPPEDGETKAPRESSSDVTK 549

DB 389 AILN-----PCHLKA----- 399
 QY 550 YQSVSAENFVENHINITO---SDKFTAKPLDINSGERNDLNRSCGVPESSASSEKAK 606
 DB 400 -----HFRARCLFELKYVAAL-----ECLDDEFCK 426
 QY 607 EPESTQOTSE-----SATNENNTNPEPOFQTEATGPSAHEETSTRDSALQDTOSDD 659
 DB 427 FPEQAHSSACDALGRDITALLFSKSDGEKKAAAGGGGPVRLRSTSRKDSI-----SED 480
 QY 660 DPEYLIGARRAAGBGRASVARIQEFFRRRKEKKEDELDTLNIRRLVAMVYKGHNS 719
 DB 481 EMVL-----RESDYQFR-----YCGHNT 501
 QY 720 RTMIKEANFEGAN--FYWSSGSDGHIPIWDRHTAEHLLEADNHVYVNCLOPHFPDPIIA 777
 DB 502 TTDIKENFNGSNAQIVYSSGSDGSFFIWEKETINILRVLOGDSIYVNCLOPHRSYCLA 561
 QY 778 SSGIDVDIKTWSPLEES 794
 DB 562 TSGIDPVYRLMNPRES 578
 RESULT 23
 AA080471
 ID AAU80471 standard; Protein; 616 AA.
 XX
 AC AAU80471;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Mouse adipose protein, adp. #2.
 XX
 KW Adipose protein; adp; obesity; transgenic animal; obesity;
 KW adipositas; bulimia; wasting; cachexia; eating disorder;
 KW body weight disorder; weight loss; cancer; infectious disease;
 KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;
 KW hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;
 KW gastrointestinal diseases; inflammatory bowel disease;
 KW ulcerative colitis; anorexia nervosa; glycogen storage disease;
 KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;
 KW infertility; acquired immunodeficiency syndrome; AIDS.
 XX
 OS Mus musculus.
 XX
 PN W0200196371-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 13-JUN-2001; 2001WO-EP06713.
 XX
 PR 16-JUN-2000; 2000US-211914P.
 PR 23-JUN-2000; 2000EP-0113049.
 PR 28-JUN-2000; 2000US-214518P.
 PR 17-APR-2001; 2001EP-0109537.
 XX
 PA (DEVE-) DEVELOGEN AG.
 XX
 PI Broenner G, Ciossek T, Dohmann C, Haeder T, Rothe M;
 XX WPI; 2002-106464/14.
 DR N-PSDB; ABK16384.
 XX
 PT Novel nucleic acid encoding adipose polypeptide which regulates, causes
 PT or contributes to obesity, useful for treating obesity, heart disease,
 PT hypertension, infertility, and controlling weight loss in cancer
 PT patients -
 PS Claim 1; Page 175-177; 188pp; English.
 CC The invention relates to a nucleic acid encoding a adipose (ADP)
 CC polypeptide which regulates, causes or contributes to obesity in an

CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,
 CC modulators of adp activity, adp antisense nucleic acids, expression
 CC vectors, adp transgenic animals are useful in the diagnosis and
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating
 CC disorders and/or disorders of body weight/body mass, weight loss due to
 CC cancer or infectious diseases, genetic disorders associated with
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative
 CC colitis, and anorexia nervosa. They are also useful for treating
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid
 CC storage diseases and for treating lipomas, and/or liposarcomas. The
 CC compositions are also useful for treating heart disease, hypertension,
 CC and infertility and for treating conditions associated with under weight
 CC e.g. enhancing or controlling fertility, controlling weight loss in
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The
 CC present sequence is an adp protein.

XX Sequence 616 AA;

Query Match 9.7%; Score 441; DB 23; Length 616;

Best Local Similarity 21.9%; Pred. No. 1.6e-23;

Matches 191; Conservative 111; Mismatches 275; Indels 294; Gaps 30;

13 DYKRSLSGLDEPSRLSRSLRGLRRETIQRLKLAATLVHDGCVNTICWDTGYILSGSD 72
 9 DLIRQVKEKGLSFERRYHVTDPFIRRLGLEAELOGSHSGVCNCEMNEKCDLLASGSD 68
 73 TKLVSNPYSRKVLVTIRSGRANFSAKFLPCTNDKOVSGSGGCVIYTVNEDEAEIN 132
 69 QHTIWDPLHKKLSMTGHTANIFSVKFLPHAGGRILLITGAADSKVHDLTYKETIH 128
 133 ROCQFCHYGTYTEIMTVPNDFELSGEDGVWF-----DTRIKTSCTEKEDKDL 187
 129 ---MGDHTNRVKRIATAPMNTFWSAEDGLIRKSMQSPAGVHTCDQ----- 178
 188 INCRRAASVAICPPPIYTLAVGCSDSVRYTRDRRLGTRANGTAGCTGMVAFITIS 247
 179 -----KPLP-----DGAQVY-----VAGHLFV 196
 248 HLNKSCKRY-----TSLCYSEDOELIVSSDYTYLPDPKODTARELTPSAERREEL 302
 197 KLPTDINSRLRVATYVTFSPNGTELVNMGEOYTLF-----LTY 238
 303 ROPVVKRLRLNGDWDGTPRARERERERERERERERERERERERERERERERERER 362
 239 KORYPTFL-----PR-KCHSEVQNGKMTN-----GVSNQVSGHLHS----- 278
 363 NRGERSRPRGTSOSDSTLTPTVSSPDLF-VSETAMEVDTPAEQFLQPTSSITMSAOA 421
 279 ---NGFRLEP-----SKGCIS--POVELPPLYERVKOQANEA-PACQW-----TQAIOL 322
 422 HSTSPTSPHSTPLTSPDSBOQSVASGHTTHQSDNNNEKLSPKPGTEPYLSLHY 481
 323 YS-QAVQKAPHNMLY-----NRAAYATKRWKWDG-----HY 354
 482 STEGTTSTIKLFTDEWSSIASSSRGIGSCKSGQESFVPOSSVOPPEGDSKPAE 541
 355 DALRDLKALISLN-----PCHLKA----- 373
 542 ESSEVYTKQEGVSAENVENHINITQ---SPKFTAKPLDSNGSERNDLNDRCGVPPE 598
 374 -----HFRILARCLFELKYVAEAL-----E 392
 599 SASSEKAKEPETSDOTSTE-----SATNENNTNPEPOFTEATGPAHSEFTSRDASL 651
 333 CLDPEKGFPEQASHSACDALGRDITALLFSKSDGEKKAAGGGGAPVRLRSTSRDST- 451
 652 QDTPDSDDDPVLLPGARYRAGPGRRSARVARIQEFFRRKKEKKEBELDTLIRPLVM 711
 452 -----SEDEMYL-----RESYVYQGR----- 468
 712 YKGRHSRTYMIKEANFNGAN--FVMSGSDCGHIFITWRHTAHLMLLEADNHNVCLOP 769

DB 469 -CGHCNTTIDIKAEANFESNMQYIVSGSDSGEFFEWEKRETNLVVLGDESIVNCIP 527
 QY 770 HEPDPLASSGIDYDIKIKSPLEESRIRNRKLADEVITNEMLETRNTIYVPAFMR 829
 DB 528 HSYCYLATSGIDPVVRLNPNPRESEDLTGRAVEDMEGASQ-----ANORRNAPLEA 581
 QY 830 MLASLNHIRADRLLEGDRSGSGOENENDEE 860
 DB 582 MLMDGY-----RITGLSSGAGASDDEDSAE 608

RESULT 24

ABB27840

ABB27840 standard; Peptide; 86 AA.

AC ABB27840;

DT 01-FEB-2002 (first entry)

DE Human peptide #491 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;

KW disease; cancer.

XX Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-496933/54.

PS New spatially-addressable set of single exon nucleic acid probes,

PT useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 10808; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

XX nucleic acid probes for measuring gene expression in a sample derived

XX from human breast and BT 474 cells. The method involves contacting

XX the probes with a collection of detectably labeled nucleic acids

XX derived from mRNA of human breast, and then measuring the label

XX bound to each probe of the microarray. The probes are useful for

XX verifying the expression of regions of genomic DNA predicted to

XX encode proteins. They are useful for gene discovery, and for

XX determining predisposition and/or prognosing breast disease. Gene

XX expression analysis is useful for assessing the toxicity of chemical

XX agents on cells. The microarray of this invention presents a far greater

XX diversity of probes for measuring gene expression, with far less bias

XX than expressed sequence tag microarrays. The method is suitable for

XX rapid production of functional information from genomic sequence. The

XX present sequence is a peptide encoded by a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from Wipo at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 86 AA;
 Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 433
 DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 60

OY 434 TPLISSPDEQRQSVASGHHTHQS 459
 DB 61 TPLISSPDEQRQSVASGHHTHQS 86

RESULT 25
 ABB33011
 ID ABB33011 standard; Peptide; 86 AA.
 AC ABB33011;
 XX 04-FEB-2002 (first entry)
 DE Peptide #517 encoded by human foetal liver single exon probe.
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 OS Homo sapiens.
 XX MO200157277-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483447/52.
 XX The present invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535 ABA1305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC Sequence 86 AA;
 SQ

Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 433
 DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 60

OY 434 TPLISSPDEQRQSVASGHHTHQS 459
 DB 61 TPLISSPDEQRQSVASGHHTHQS 86

RESULT 26
 ABB18484
 ID ABB18484 standard; Protein; 86 AA.
 AC ABB18484;
 XX 23-JAN-2002 (first entry)
 DE Protein #483 encoded by probe for measuring heart cell gene expression.
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 OS Homo sapiens.
 XX MO200157274-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00666.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 CC hearts -
 CC Claim 15; SEQ ID NO 20254; 530pp; English.
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535 ABA1305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC Sequence 86 AA;
 SQ

Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 433
 DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSAQAHSTSSPTESPHS 60

OY 434 TPLISSPDEQRQSVASGHHTHQS 459
 DB 61 TPLISSPDEQRQSVASGHHTHQS 86

RESULT 27
 AAM53809
 ID AAM53809 standard; Protein; 86 AA.
 XX
 AC AAM53809;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 25914.
 XX
 KW Human: brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 PS Example 4; SEQ ID NO: 25914; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 CC
 SQ Sequence 86 AA;
 XX
 Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 374 GTSQSDISTLPVPSSPDLEVTAMEVDPAPQFLQPSSTMSQAASHSTSPSPHS 433
 DB 1 GTSQSDISTLPVPSSPDLEVTAMEVDPAPQFLQPSSTMSQAASHSTSPSPHS 60
 OY 434 TPLSSPDSEQRQSVASGHTTHQS 459
 DB 61 TPLSSPDSEQRQSVASGHTTHQS 86
 RESULT 28
 AAM6195
 ID AAM6195 standard; Protein; 86 AA.
 XX
 AC AAM6195;
 XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26501.
 XX
 KW Human: bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 26501; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 CC
 SQ Sequence 86 AA;
 XX
 Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 374 GTSQSDISTLPVPSSPDLEVTAMEVDPAPQFLQPSSTMSQAASHSTSPSPHS 433
 DB 1 GTSQSDISTLPVPSSPDLEVTAMEVDPAPQFLQPSSTMSQAASHSTSPSPHS 60
 OY 434 TPLSSPDSEQRQSVASGHTTHQS 459
 DB 61 TPLSSPDSEQRQSVASGHTTHQS 86
 RESULT 29
 AAM14066
 ID AAM14066 standard; Protein; 86 AA.
 XX
 AC AAM14066;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #500 encoded by probe for measuring cervical gene expression.
 DE Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 KW
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID NO 18892; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs; see A110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ffp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 86 AA;
XX
Query Match 9.7%; Score 437; DB 22; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.5e-24;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 GTSQSDISTLPVSPSPDLEVSPTAMEVDYPAEOPLOPSTSTMSQAHSSTSPRESPHS 433
DB 1 GTSQSDISTLPVSPSPDLEVSPTAMEVDYPAEOPLOPSTSTMSQAHSSTSPRESPHS 60
XX
QY 434 TPLISSPDSEQRQSVASGHHTHQS 459
DB 61 TPLISSPDSEQRQSVASGHHTHQS 86
XX
RESULT 30
AAM26472
ID AAM26472 standard; Protein; 86 AA.
XX
XX AAM26472;
AC
XX
XX 17-OCT-2001 (first entry)
DE
XX Peptide #509 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX WO200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR

PR 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID NO 26741; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs;
CC see A113135-A115746). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
SQ Sequence 86 AA;
XX
Query Match 9.7%; Score 437; DB 22; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.5e-24;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 GTSQSDISTLPVSPSPDLEVSPTAMEVDYPAEOPLOPSTSTMSQAHSSTSPRESPHS 433
DB 1 GTSQSDISTLPVSPSPDLEVSPTAMEVDYPAEOPLOPSTSTMSQAHSSTSPRESPHS 60
XX
QY 434 TPLISSPDSEQRQSVASGHHTHQS 459
DB 61 TPLISSPDSEQRQSVASGHHTHQS 86
XX
RESULT 31
AAM01806
ID AAM01806 standard; Protein; 86 AA.
XX
XX AAM01806;
AC
XX
XX 09-OCT-2001 (first entry)
DE
XX Peptide #488 encoded by probe for measuring human breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
OS
XX
XX WO200157270-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US00661.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX

PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 XX
 PS Claim 27; SEQ ID No 10546; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see A1100010-A1110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridizes at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 86 AA;
 XX
 Query Match 9.7%; Score 437; DB 22; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 374 GTSQSDISTLPTVPSSPDLEVSETAEMVDTPAEOFLQPSSTSMQAHSSTSPSPHS 433
 DB 1 GTSQSDISTLPTVPSSPDLEVSETAEMVDTPAEOFLQPSSTSMQAHSSTSPSPHS 60
 OY 434 TPLSSPSEQRQSVESAGHHTHHOS 459
 DB 61 TPLSSPSEQRQSVESAGHHTHHOS 86
 XX
 RESULT 32
 ABG35845
 ID ABG35845 standard; Peptide; 86 AA.
 XX
 AC ABG35845;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 25510.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberculous scleriosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kartagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
 DR WPI: 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 PS Claim 27; SEQ ID No 25510; 634pp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberculous scleriosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 86 AA;
 XX
 Query Match 9.7%; Score 437; DB 23; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 374 GTSQSDISTLPTVPSSPDLEVSETAEMVDTPAEOFLQPSSTSMQAHSSTSPSPHS 433
 DB 1 GTSQSDISTLPTVPSSPDLEVSETAEMVDTPAEOFLQPSSTSMQAHSSTSPSPHS 60
 OY 434 TPLSSPSEQRQSVESAGHHTHHOS 459
 DB 61 TPLSSPSEQRQSVESAGHHTHHOS 86
 XX
 RESULT 33
 AAG18387
 ID AAG18387 standard; Protein; 471 AA.
 XX
 AC AAG18387;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 19771.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX EPI03405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147182.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149398.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152263.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157173.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159337.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 9.6%; Score 436.5; DB 21; Length 471;

Best Local Similarity 18.7%; Pred. No. 2,2e-23; Matches 147; Conservative 66; Mismatches 153; Indels 419; Gaps 12;

QY 9 HLLMDVRRKSLGLEDPSRLSRYLGRREFIQRKLEATLVNVDGCVNTICWMDTGEYILS 68
 DB 13 HSYVAVWERELGILNRSFNSASSEDLLRLGLDKLDRKKGCVNTVSPFADGDILIS 72
 QY 69 GSDPKIVISNDYRKRYLTTRSGHRANIPSAKFLPTNDKOIVSCSGDVIYFTNVED 128
 DB 73 GSDRQVILMWMQVATSVLSPDSGFHNNIFQAKFMPFSDRITVNSADKQVRSKILS 132
 QY 129 AETNQCQCFCHYCTTVELMYPNDPYTFELSCGEGDGYRWPTRIKIKTSCTKDKDDILL 188
 DB 133 GQVEVSL-LGKHQGVNHLAVPEPSFSPFYTCGEDGAVKHFDLRTVRVATNLETC-EAKF 190
 QY 189 NCRRAATVATCPPIPYLLAVGCSDSVRIYDRMLGTRATGNVGRGTGVARFIPSH 248
 DB 191 NLVYVLIHALADPNRPGLLAVAGNDEYARVDIRYRESCWTF-----TOPIDHFCGH 245
 QY 249 L-NKSCSVTSLCYSEDQELIVSYSDYIYLPDKDDTJAEKLPKSAEERREELROPV 307
 DB 246 LIGDDHVGITGLAFS-DSSELLASYSDEFIYTFP----- 279
 QY 308 KRLRLGDMWSTGPRAPRESEREDGEGSPNVSLMQKMSDMLSRWFEEASEVAQSNRGRG 367

DB 280 -----DM-----GLG 284
 QY 368 RSRPRGTSOSDISTLYTPVSSPDLEVSETAMEVDTPAEQFLQPSSTSTMSQAQAHSTSP 427
 DB 285 -----P 285
 QY 428 TESPHSTPLSSPDSPQROSVASGHTTHQSDNNNEKLSPKGTGPVLSLHYSTGTT 487
 DB 286 TPYSSSTK----- 293
 QY 488 TSTIKLFTDMSVSIASSSGISGSHCKSEGOEESFYPOSSVQPPGDSFTKAPDESEEDV 547
 DB 294 -----TERMPPO----- 301
 QY 548 TKYQGVSAENPVENHINTQSDKFTAKPLDINSGERNDLINDRSCGVPEASSEKAKE 607
 DB 302 ----- 301
 QY 608 PETSDQTESATNENNTNPEQFQTEATGSAHEETSTRDSALQDTDDSDPVLIPGA 667
 DB 302 ----- 301
 QY 668 RYRAGPDRRSAYARIQEEFRKRRKEMEBELDTNIRPLVMMYKGRHSRTMIKEAN 727
 DB 302 -----YKEHTNRET-YKGVN 316
 QY 728 FNG--ANFYMSGSDGHIITWDRHTAEHMLLEADNHYVNCLOPHFPDPLASSGIDYDI 785
 DB 317 FPGKCEYVAGSDGCRIFIRWKRKDELRLAMADRHVYNCIESHPHMLCMSSGIDTDI 376
 QY 786 KIWSF 790
 DB 377 KIWTF 381

RESULT 34
AAG50441

ID AAG50441 standard; Protein; 471 AA.

XX AAG50441;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 63925.

KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130049.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132488.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149992.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153788.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157717.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158368.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

```

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 9.6%; Score 436.5; DB 21; Length 471;
 Best Local Similarity 18.7%; Pred. No. 2.2e-23;
 Matches 147; Conservative 66; Mismatches 153; Indels 419; Gaps 12;

```

QY 9 HILMDVYKRSLSGLIEDPSRLSRVLRGRREFLORLKLATLVNHDGCVNTICMNDTGEYIIS 68
DB 13 HAVVNWVWERELGLPRRSNRFSSASEDLRLGLDKLDRKRGCVNTVSFNADGDIILS 72
QY 69 GSDDTKLVISNPKRVLTIRSGHRANIFSAKFLPCTNDKQIVSCSGGVIFYTVNEOD 128
DB 73 GSDDBQVILMDQVATSVKLSFDSGHNIFQAKFMFSDRTIVTSAADKQVRSKILMS 132
QY 129 AETNRQCOFTCHGYTTEIMTVPNDDPYTLSCGEGTVAWMPRIKTSCTKECKDDIIL 188
DB 133 GOVERSL-LGRKQGVYHKAVERGSPFSTYTGEDGAVAHFDRKVAATNLFCK-EAKF 190
QY 189 NCRRAATVAICPIPIYVLAVGCSDSVRIYDRMLGTRATGNVGRGTGAVARFIPSH 248
DB 191 NLVVYLAIAVDPRMDGLAVAGMDEYARVDIRYSRSGWNF-----TOPIDHRCPEH 245
QY 249 L-NMNSCAVTSICSEDOELIVSSDIYIFLPRKDPDARLKTSPASERREELROPV 307
DB 246 LIGDHVIGITGLAFS-DSSELLASYSDEFIYLFTP----- 279
QY 308 KRLRLRGDMSDTPRPARPESERERDGEOSPVSIMQMSDMLSRWFEASEVAQSNRG 367
DB 280 -----DM-----GLG 284
QY 368 RSRPRGTSOSDISTLPTVPSSPDLEVESETAMEVDTPAEOLFOPSTSTMSQAHSSTSP 427
DB 285 -----P 285
QY 428 TESPSPPLSSPDSEQRQSVASGHNTHQSDNNNEKLSKPRGCEVYLSLHSTEGTT 487
DB 286 TPYSSTT----- 293
QY 488 TSTIKLFTDEMSSIASSSRGISGHSCKSEGOEESFVPOSSVOPPEGDETKAPERSSEV 547
DB 294 -----TEEMTPQ----- 301
QY 548 TKYQEGVSAENPVENHINITOSDKFTAKPLDLSNGERNLMDLNSCGVPEESASSEAKE 607
DB 302 ----- 301
QY 608 PETSDQTSSTASATNENNTNPEPQTEATGPSAHEETSTRSALQDDDDSDDDVILPGA 667
DB 302 ----- 301
QY 668 RYRAGPDRSAAVARIOEFFRRKREKEMELDTLIRPLVYKMGHRSRTMIKEAN 727
DB 302 -----VYKEHNREI-VAGVN 316
QY 728 FWG--ANFVWSGSDCGHIFIMDRHTAEHLMLLEADNHVVCLOPHPPDILLASSGDIYDI 785

```

```

DB 317 FFGPKCEVWVSGSDCGRIFTRKRDGELRAMEDRHHVNCIESHPHPLMCSGIDTID 376
QY 786 KIMSP 790
DB 377 KIWP 381

```

RESULT 35

AA85870 standard; Protein; 816 AA.

AA85870;

13-SEP-1996 (first entry)

WD-40 domain-contg. Mus musculus protein.

WD40 repeat region; beta-transducin; protein-protein interaction; drug; intracellular signaling; protein kinase C; homology; motif; modulator; receptors of activated protein kinase; enzyme activity; isozyme; human.

Mus musculus.

WO9521252-A2.

10-AUG-1995.

31-JAN-1995; 95WO-US01210.

01-FEB-1994; 94US-0190802.

(STRD) UNIV IELAND STANFORD JUNIOR.

Mochly-Rosen D, Ron D;

WPI; 1995-283772/37.

New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the activity of a protein, eg. protein kinase C, which interacts with a protein contg. a WD-40 region.

Example 5; Page 135-138; 351pp; English.

proteins AAR85851-92 are protein which contain at least one WD-40 (also called beta-transducin homologous) amino acid repeat motifs. The WD-40 regions are involved in protein-protein interactions between proteins involved in intracellular signaling. An example of such an interaction is between protein kinase C and receptors of activated protein kinase (RACK), esp. RACK-1 (AAR85850). Proteins AAR85851-82 were isolated based on homology with beta-transducin, whereas proteins AAR85882-92 were isolated based on homology with the WD-40 consensus sequence (AAR85893). The proteins were used to construct the peptides AAR84928-R85063 and AAR85786-R85842. The peptides can be used to identify target proteins contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of proteins involved in protein-protein interaction and to screen for drugs that will affect protein-protein interaction involving WD-40 domains.

Sequence 816 AA;

Query Match 9.5%; Score 429.5; DB 16; Length 816;
 Best Local Similarity 18.7%; Pred. No. 1.7e-22;
 Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

```

QY 8 PHILMDV-----RKRSLGLIEDPSRLSRVLRGRREFLORLKLATLVNHDGCVNTICMNDTG 63
DB 377 PRPRMNVLSALRDQGLG--SSGRFYVEACGAKLVQNFSLHVEGSGCVNTVHFNQHG 434
QY 64 EYILSGSDTKLVISNPKRVLTIRSGHRANIFSAKFLPCTNDKQIVSCSGGVIFYTV 123
DB 435 TLIASGSDDLKVIYVMDLKRKSVLNFDSGHNITLQAKFLPNCNDAILAMGRCQYARA 494
QY 124 NVEQDAETNRQCOFTCHGYTTEIMTVPNDDPYTLSCGEGTVAWMPRIKTSCTKECK 183

```



```

Db 495 QLSAVACTHTKRLVKGASHRLGLEPDSFRRLTSEDVAVFEDIDLRQAHPAK----- 550
Qy 184 DDILI-----NCRRAATSAICPPIPYLAVGSSSVRIYDRMLGTRATGVAAGRTTG 239
Db 551 --LLVIDGKRVGLTYFVVPANVYGFVAGGDDFMRIYDORAKIDENV-----NNG 600
Qy 240 MVAFTISHL--NNSCRVTSICYSDEQGLVYSSDYIYLFDPKDDPARELKTPSAE 297
Db 601 VLKFCFPHLLSDYPAPHITSLMYSDDETLASINDEDIYF----- 643
Qy 298 RRELQDPYVKRLRLRGDMDTPRARPERERDEQSPVNSLMQMSDMLRWEAS 357
Db 644 ----- 643
Qy 358 EVAQSNRGRSRPRGTSQSDISTLPLVPSPDLEVSETAMEVDTPAQFIQPSITSTM 417
Db 644 ----- 643
Qy 418 SQAQHTSPTSPHSTPLLLSPDSEQSVASGHHTHQSDNNNEKLSPKGTGPVL 477
Db 644 ----- 643
Qy 478 SLHSTEGTTSTIKLNTDEWSSIASSRGIGSHCKSEGOESFVQSSVOPPEGDSET 537
Db 644 ----- 643
Qy 538 KAPESSBDVTKYQBGVSAENPVENHINITQSDKPTAKPLDSSGERNDLNDRCGYPE 597
Db 644 ----- 643
Qy 598 ESASSEKAKEPETSQSTESATNENNTPEQFQTEATGSAHESTSPDSALQDSTD 657
Db 644 ----- 648
Qy 658 DDDPVLIPGARYRAGGDRRAVARIQEFFFFRRRKEKMEELDTLNRPLVMKYNGHR 717
Db 649 D-----GAQY-----AKR-----YNGHR 661
Qy 718 NERTMIKEANFWG--ANFVMSGSDGHIPTDRHTAHLMLLEAD-NHVVNCLQPHFDP 774
Db 662 NNST-VKGVYFYGPKEEFVMSGDCGHITWEKSSCOIYQFLDEDEGTINCIDSHYLP 720
Qy 775 ILASSGIDYDIKISPLESRIFNRKLA--DEVITRNELMLEE--TNTITVPAS---FM 827
Db 721 VLASSGLDHEVKIWSPIAEP--SKLAGLKNVYIKIKLRKNDFTLRTSLFNNSMLCFL 777
Qy 828 LRMLASLNH-----IRADRLGEDRSQSGQENENED 859
Db 778 MSHVYQSNYGRSWRGIRINAGGDFDSSSSSEETNOE 815

RESULT 36
ABG27224
ID ABG27224 standard; Protein; 633 AA.
XX
AC ABG27224;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27215.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX

```

```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT.
PI WPI; 2001-639362/3.
XX N-PSDB; AAS91411.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 57583; 103pp; English.
XX

```

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensic, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG00010-ABG30377 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 633 AA;

Query Match 8.9%; Score 404; DB 22; Length 633;

Best Local Similarity 17.9%; Pred. No. 8.5e-21;

Matches 160; Conservative 81; Mismatches 168; Indels 486; Gaps 21;

```

Qy 2 SRGGSTPHLLMDY-----KRSGLLEPDSRLRSRY--LGRREFLQRLKLEATLVNHDGCYN 55
Db 179 SETSALPRPMQVYVTLHQRDLG---SRREYVACGARAFOREFLQRLADHVCYN 234
Qy 56 TICMDNTGEYILSGSDTKLVISNPYSRKVLITRSGRANIFSARFELPCTNDKOIVSCS 115
Db 235 TVHFNQGRTRILASSGDDKIVYVMDWRQRYVLPFESGHTNNVQAKPLPNCGSTLAMCA 294
Qy 116 GDGVIFYTVNEQDAETNRQOCFTCHYGTIYELMTVPNDPTFTLSCGEDGYRWMDIR--- 172
Db 295 RDGQVAVAEILNASYFNNTKCAQHNGRPAKRLLEPDSPKYFLTSGDAVFTIDLRQDR 354
Qy 173 --IKTSCTEKEDCKDDILINORRAATSAICPPIPYLAVGSSSVRIYDRMLGTRATG 230
Db 355 PASKVAVTRENDR-----KVLGYITTVAPANTYQFVAVAGGDDFMRIYDORAKIDKE-- 405
Qy 231 NYAGRGTTGVAVRIFPSHLNKKSC---RYSLSYSEDOGLVYSSDYIYLFDPKDDT 286
Db 406 -----NNGVLKKTFTPHLVN--CDFPTNITCVVYSHDGTLLASYNDDDIYLFN----- 452
Qy 287 ARELKTPSAEERREELQDPYVKRLRLRGDMDTPRARPERERDEQSPVNSLMQMS 346
Db 453 ----- 452
Qy 347 DMLSRWFEASVAQSNRGRSRPRGTSQSDISTLPLVPSPDLEVSETAMEVDTPAE 406
Db 453 ----- 452

```

QY 407 QELPSTSTMSQAHSSTSPESPHSTPLSSPSEQRQSVASGHTTHQSDNNNEKL 466
 Db 453 ----- 452
 QY 467 SPKPGTGPVLIAHSTGTTSTIKLNPTEDEMSIASSRNGISGHSCKSEGESEFVPOS 526
 Db 453 ----- 452
 QY 527 SVQPPGDESETKAPESEEDVTKOEGVSAENPVENHINIOSDKFTAKPLDSNGERN 586
 Db 453 ----- 452
 QY 587 LINDRCGVPEASASEKAKPETSDQSTESATNENTNPEPOQTATGSAHEETST 646
 Db 453 ----- 457
 QY 647 RDSALQDTDDSDPVLIPGARYRAGPGRRSAAVARIOEPPRRRERKEMELDTLNR 706
 Db 458 ----- 464
 QY 707 PLVKNVKGHRNRTMKRANFNG--ANFVMSGSDCGHIFWDRHTAEHLMLLEADNH-V 763
 Db 465 -----FKGHRNN--TTKGVNFTGPRSEFVVSQDCHIFEWKSSCOIIQFLKGRRG 517
 QY 764 VNCIQHPPEPIIASGIDYDIKITSPLIESRIFNRKLADP-----VITREML--EET 816
 Db 518 INCLEPHPLPYLACSGLDHDKIWP-----TAKAPTELGLKVKYIKKKMERDEDS 570
 QY 817 RNTTVASFMLRLASLNH-----RADRLP--GDRSGSGGENEDEE 860
 Db 571 LHHGSLFDQYMLWFL--LRHYTORGRHQDWKSGEAFPPDESDSSSTSESEE 623
 RESULT 37
 AAU80461
 ID AAU80461 standard; Protein; 385 AA.
 AC AAU80461;
 DT 12-MAR-2002 (first entry)
 DE Fruit fly adipose protein, adp, #1.
 XX
 KW Adipose protein; adp; obesity; transgenic animal; obesity;
 KW adipositas; bulimia; wasting; cachexia; eating disorder;
 KW body weight disorder; weight loss; cancer; infectious disease;
 KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;
 KW gastrointestinal diseases; Cushing's syndrome; endocrine disorder;
 KW ulcerative colitis; anorexia nervosa; glycogen storage disease;
 KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;
 KW infertility; acquired immunodeficiency syndrome; AIDS.
 OS Drosophila melanogaster.
 XX
 PN WO200196371-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 13-JUN-2001; 2001WO-EP06713.
 XX
 PR 16-JUN-2000; 2000US-211914P.
 PR 23-JUN-2000; 2000EP-0113049.
 PR 28-JUN-2000; 2000US-214518P.
 PR 17-APR-2001; 2001EP-0109537.
 XX
 PA (DEVE-) DEVELOGEN AG.
 XX
 PI Broenner G, Ciossek T, Dohmann C, Haeder T, Rothe M;
 DR WPI; 2002-106464/14.
 DR N-PSDB; ABK16344.
 XX

PT Novel nucleic acid encoding adipose polypeptide which regulates, causes
 PT or contributes to obesity, useful for treating obesity, heart disease,
 PT hypertension, infertility, and controlling weight loss in cancer
 PS patients.
 PS
 PS Claim 1; Page 138-139; 188pp; English.

CC The invention relates to a nucleic acid encoding a adipose (ADP)
 CC polypeptide which regulates, causes or contributes to obesity in an
 CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,
 CC vectors, adp transgenic animals are useful in the diagnosis and
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating
 CC disorders and/or disorders of body weight/body mass, weight loss due to
 CC cancer or infectious diseases, genetic disorders associated with
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative
 CC colitis, and anorexia nervosa. They are also useful for treating
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid
 CC storage diseases and for treating lipomas, and/or liposarcomas. The
 CC compositions are also useful for treating heart disease, hypertension,
 CC and infertility and for treating conditions associated with under weight
 CC e.g. enhancing or controlling fertility, controlling weight loss in
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The
 CC present sequence is an adp protein.

Sequence 385 AA;

Query Match 8.88; Score 399; DB 23; Length 385;
 Best Local Similarity 32.6%; Pred. No. 9.3e-21;
 Matches 93; Conservative 50; Mismatches 120; Indels 22; Gaps 5;

QY 10 LMDVRRKSLGLDPSLRSGRYLGRREFIQRLKLEALNVHDCVNTTCWMDGTYLISG 69
 Db 28 LHM--QRQYGHLEDEMLRRLASPAYVDRLQEAALVHGEGVNCLEVTTOGMILASG 85
 QY 70 SDDTKLVISNPSRKVLTTRSGHRANIFSAKFLPCTNDKQIVSGSGGVIFTNVQDA 129
 Db 86 SDDYRWIMDPFRKKLVHVRITKHLGVFVKLPKNNISIVATCADKTIYVDINDPN 145
 QY 130 ETNRQCFCHYGTTEIKMVPNDPTFLTSCGSDGVYRWMDTRIKTSCTEKDCKDILIN 189
 Db 146 ETLPSC--ICHFSRANRLAADSPPVFWGAGDGLQIDIREPHRCRREGIGVRLIN 203
 QY 190 C-----RRAATSVATCPPIPYLLAVGCSDSVRYDRMIGTRATGNYAGRTTGMYARF 244
 Db 204 LHDQLENTAKCLAINPRTREYLAAGTNDPFARVYDRKKLPST-----NNGISACVAYY 258
 QY 245 IPSHLNKSKCR-----VTSICYSDEQGLIVSYSDYIYLF 281
 Db 259 APQIVKNISRNIVHEPRGITVLTENGCTELLVNICGEHYVRF 303

RESULT 38
 ABB57847
 ID ABB57847 standard; Protein; 748 AA.
 XX
 AC ABB57847;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 333.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX

PF 23-MAR-2001: 2001WO-US09231.
 XX 23-MAR-2000: 2000US-191637P.
 PR 11-JUL-2000: 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI: 2001-656860/75.
 DR N-PSDB; ABL01950.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 333; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB161737-AB172072).
 CC (AB161737-AB172072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 748 AA:
 XX
 Query Match 8.2%; Score 373; DB 22; Length 748;
 Best Local Similarity 17.3%; Pred. No. 2.1e-18;
 Matches 147; Conservative 73; Mismatches 163; Indels 468; Gaps 14;
 QY 31 YLGRREFIQLKLEATNVAGCVNTICWMDTGEYILSGSDTKLVISNPKYKLVLTIR 90
 DB 306 YYSROVVEQMTLLSLNHHGCVNSLNFNRAAGLCSGSDLLTVMDWMEKOLHFR 365
 QY 91 SGRANIFSAKFLPCTNDKOTVSCSGDGQVIFYTNEQDAETNRQCFCHYGTTEIMV 150
 DB 366 SGHNMNFQFKFIDSACGLDIVSSRDQVRSYVPPSGVYKPRLYTHSESVKILTV 425
 QY 151 PNDYTLSCGEGDGVKMFPRITKSCCKEDOKDILINC-----RRATSVATCP 201
 DB 426 PHSRHEHLSAGEDAAYVHFELASNAAT-----TMRVCYNDENRERVALFSAHP 478
 QY 202 PIPYLLAVGSDSSVRIYDRMLGTRATGNYAGRGTTGNVAFIPSHLNKSCRYTSLCY 261
 DB 479 YAFPCVSG-SDDLIRVYDKRL-AKAIHOMAPRNL-----LEAOITQITCAVY 525
 QY 262 SEDQELIVSYSSDIYIFDPKODTARELKTPSAEFRRELKOPPYKRLRLRGDWDGTP 321
 DB 526 NMSGSEILASYSDAGIYLFDSRNYN-----RDY----- 554
 QY 322 RARPERERDEGEQSPNVSLQMDSMDLSRWEFEASEVAQSNRGSRPRGGTSGSDIS 381
 DB 555 ----- 554
 QY 382 TLTPVSPDLEVSETAMEVDTPAEOFTLPSTSTMSAQAHSTSPESHSTPLSSPD 441
 DB 555 ----- 554
 QY 442 SGROSVASGHHHTHQSNNNEKLSPPKPTGEPVLSLHSTGTGTTSTIKLNFTEWSS 501
 DB 555 ----- 554
 QY 502 IASSRGIGSHCKSEQGESFVPOSSVOPPEGSETKAPESSSDYTKYQGSVAENPYE 561
 DB 555 -----LHC----- 557
 QY 562 NHINTQSKFTAKPLDLSNGERNDLDRSCGVPEESASEKAKETSDOTSTESATN 621

DB 558 ----- 557
 QY 622 ENNTNPEPQTEATGSAHEHSTSTRDSALQDTPDDSDPVLIPGARYRAGPGRRSAYA 681
 DB 558 ----- 557
 QY 682 RIQEFFRRKREKEMEDLTINIRPLVKMYKGRNSRIMKEANFWG--ANTVMSSGD 739
 DB 558 -----YKGHINSRT-IKGVNFPGRSEYIVSGSD 585
 QY 740 CGHIFWDRHTAEHMLLEADN-HVYNCLOPHPDPLIASSGIDYDKIMSPLESRIFN 798
 DB 586 CGNIFWMDKTEALINMYKGDHVGVCNLEPHMPMVLATISGLEADVAKIWPQGERKLP 645
 QY 799 RKLADVITRN-----ELMLEETRNIT-----TVPASFMLMLASLNHTRADRL 843
 DB 646 EDLLKQTLQKRNFRNIVASIDLDINNFYIFRIGFQSGASHLRQRTHTSLGHOMKDN-S 704
 QY 844 GDRSESGOEN 854
 DB 705 SRSNGSNASN 715
 RESULT 39
 ABB27841
 ID ABB27841 standard; Peptide; 62 AA.
 XX
 AC ABB27841;
 XX
 DT 01-FEB-2002 (first entry)
 DE Human peptide #492 encoded by breast cell single exon nucleic acid probe.
 XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 KW
 OS Homo sapiens.
 XX
 PD WO200157271-A2.
 XX
 PF 09-AUG-2001.
 XX
 PR 30-JAN-2001: 2001WO-US00662.
 XX
 PR 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-496933/54.
 XX
 DR
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 27; SEQ ID NO 10809; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and pr 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 62 AA;

Query Match
 Best Local Similarity 100.0%; Score 338; DB 22; Length 62;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGGVIFYTNVEDAETNRCCOFTCHYGT 144
 DB 1 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGGVIFYTNVEDAETNRCCOFTCHYGT 60
 OY 145 YE 146
 DB 61 YE 62

RESULT 40

ABB33012
 ID ABB33012 standard; Peptide: 62 AA.

AC ABB33012;

DT 04-FEB-2002 (first entry)

DE Peptide #518 encoded by human foetal liver single exon probe.

KM Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human foetal liver -
 XX
 XX Claim 27: SEQ ID NO 25647; 639pp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 62 AA;

Query Match
 Best Local Similarity 100.0%; Score 338; DB 22; Length 62;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGGVIFYTNVEDAETNRCCOFTCHYGT 144
 DB 1 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGGVIFYTNVEDAETNRCCOFTCHYGT 60
 OY 145 YE 146
 DB 61 YE 62

Search completed: March 18, 2003, 15:34:01
 Job time : 54 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:30:15 ; Search time 37 Seconds

(without alignments)
4789.203 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRRGSRPHLMVDKRRSLG.....RLEGRSRSGSGEENEDDEE 860

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 120 summaries

Database :

SPTREMBL 21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteriaphage:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3972	87.7	876	11 09DC22	09DC22 mus musculus
2	2663	58.8	513	4 08TB19	08TB19 homo sapien
3	2647	58.5	533	4 09P000	09P000 homo sapien
4	2258	49.9	485	4 09UD05	09UD05 homo sapien
5	590.5	13.0	754	10 09LEU6	09LEU6 homo sapien
6	580	12.8	757	10 09AB03	09AB03 arabisidopsis
7	547.5	12.1	809	3 094527	094527 schizosacch
8	508.5	11.2	628	5 09V8C7	09V8C7 drosophila
9	448.5	9.9	481	10 09M1E5	09M1E5 arabisidopsis
10	443.5	9.8	597	4 012839	012839 homo sapien
11	429.5	9.5	747	11 001078	001078 mus musculus
12	405.5	9.0	493	10 09T001	09T001 arabisidopsis
13	384.5	8.5	426	10 09S2M8	09S2M8 arabisidopsis
14	373	8.2	743	5 08SX92	08SX92 drosophila
15	373	8.2	748	5 09W091	09W091 drosophila
16	330	7.3	615	4 09NV87	09NV87 homo sapien

17	249.5	5.5	1514	5 08SY55	08SY55 drosophila
18	245.5	5.4	355	4 09HEB2	09HEB2 homo sapien
19	232.5	5.1	958	4 096JK2	096JK2 homo sapien
20	213.5	4.7	747	5 09V5L7	09V5L7 drosophila
21	212.5	4.7	298	5 095TD6	095TD6 drosophila
22	211.5	4.7	747	5 09B1I5	09B1I5 drosophila
23	207.5	4.6	131	10 09FWE0	09FWE0 oryza sativa
24	204.5	4.5	1217	4 060336	060336 homo sapien
25	202	4.5	5170	5 017490	017490 caenorhabdi
26	201	4.4	2586	5 09VTK8	09VTK8 drosophila
27	200	4.4	1183	5 0960U6	0960U6 drosophila
28	200	4.4	2351	5 09VQ54	09VQ54 drosophila
29	200	4.4	6994	5 017343	017343 caenorhabdi
30	199	4.4	493	5 023330	023330 caenorhabdi
31	198	4.4	3507	5 023587	023587 caenorhabdi
32	195.5	4.3	791	13 09PGL1	09PGL1 fugu rubrip
33	195.5	4.3	1162	5 09VB40	09VB40 drosophila
34	194.5	4.3	3060	5 09VAV4	09VAV4 drosophila
35	194	4.3	1680	5 09UQA2	09UQA2 caenorhabdi
36	193	4.3	451	3 09P785	09P785 schizosacch
37	193	4.3	852	10 09S255	09S255 arabisidopsis
38	192.5	4.3	985	12 067643	067643 gallid hep
39	191.5	4.2	2951	5 09W320	09W320 drosophila
40	191	4.2	1508	11 09R0L0	09R0L0 mus musculu
41	190	4.2	788	4 096A69	096A69 homo sapien
42	189	4.2	819	4 096ST2	096ST2 homo sapien
43	189	4.2	1142	4 08RCV4	08RCV4 homo sapien
44	189	4.2	2174	4 08RAT0	08RAT0 homo sapien
45	188.5	4.2	2174	4 08GQRO	08GQRO drosophila
46	187.5	4.1	2004	16 0970P7	0970P7 streptococc
47	186.5	4.1	2283	2 08VQ99	08VQ99 staphylococ
48	186	4.1	1165	4 095819	095819 homo sapien
49	185.5	4.1	1601	5 08WSM1	08WSM1 caenorhabdi
50	184	4.1	474	5 08R1B6	08R1B6 dictyostell
51	184	4.1	1255	2 086167	086167 salmoneilla
52	184	4.1	5722	5 09VPL9	09VPL9 drosophila
53	184	4.1	5322	5 09N164	09N164 drosophila
54	183	4.0	727	5 018918	018918 caenorhabdi
55	183	4.0	2271	16 0990Y4	0990Y4 streptococc
56	183	4.0	2297	3 09HGK6	09HGK6 candida alb
57	180.5	4.0	1175	4 075172	075172 homo sapien
58	180.5	4.0	2910	10 09FND5	09FND5 arabisidopsis
59	178	3.9	1027	4 09NST7	09NST7 homo sapien
60	178	3.9	1051	10 09LXN4	09LXN4 arabisidopsis
61	177.5	3.9	1063	16 09CH86	09CH86 lactococcus
62	177.5	3.9	1100	13 057576	057576 cynops pyrr
63	176.5	3.9	802	5 019764	019764 caenorhabdi
64	176.5	3.9	5327	5 076891	076891 xenopus lae
65	176	3.9	703	13 090YE6	090YE6 xenopus lae
66	176	3.9	1459	5 017084	017084 caenorhabdi
67	176	3.9	1838	13 09DC88	09DC88 morone saxa
68	175.5	3.9	2768	5 09VC00	09VC00 drosophila
69	175.5	3.9	534	3 094317	094317 caenorhabdi
70	175	3.9	1604	3 014148	014148 schizosacch
71	175	3.9	1829	5 022248	022248 caenorhabdi
72	175	3.9	734	5 08T1A0	08T1A0 dictyostell
73	174.5	3.9	1587	10 09C8N3	09C8N3 arabisidopsis
74	174.5	3.9	734	5 091365	091365 caenorhabdi
75	174	3.8	2232	5 021822	021822 caenorhabdi
76	174	3.8	870	5 09ESC8	09ESC8 mus musculu
77	173.5	3.8	2083	5 09N435	09N435 caenorhabdi
78	173.5	3.8	695	10 09S039	09S039 caenorhabdi
79	173.5	3.8	81	173	093H4 caenorhabdi
80	173	3.8	353	5 08T776	08T776 branchiosteo
81	172.5	3.8	1233	11 014515	014515 homo sapien
82	172.5	3.8	664	4 043379	043379 homo sapien
83	172	3.8	1157	5 09VTA7	09VTA7 drosophila
84	172	3.8	87	171.5	059947 streptococc
85	171.5	3.8	4498	2 09W223	09W223 drosophila
86	171.5	3.8	819	5 08SXM2	08SXM2 drosophila
87	171	3.8			
88	171	3.8			
89	171	3.8			

Db 774 EADNHVNCVLPHPDPLASSGIDYDIKIMSPLEESRIFNRKLADYITNMLLEETR 833
 QY 818 NTIVPASFMRLMNLNHRADRLDGRSGSGQENENDEE 860
 Db 834 NTIVPASFMRLMNLNHRADRLDGRSGSGQENENDEE 876

RESULT 2

Q8TB19 PRELIMINARY; PRT; 513 AA.
 AC Q8TB19; 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PRIMARY B-CELLS;
 RA Strusberg R.;
 RL Submitted (Mar-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC025262; AAB25262.1;
 SQ SEQUENCE 513 AA; 56814 MW; 0D00CB9319C00F8 CRC64;

Query Match 58.8%; Score 2663; DB 4; Length 513;
 Best Local Similarity 99.8%; Pred. No. 2.5e-165;
 Matches 512; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 348 MLSRMEFEASEVAQSNRGRSRPGTSGDSISTLPYSSPDLEVSSTAMEVDTPAQ 407
 Db 1 MLSRMEFEASEVAQSNRGRSRPGTSGDSISTLPYSSPDLEVSSTAMEVDTPAQ 60
 QY 408 FLQPTSTMSQAASHSTSPSTPLSSPDSEQOSVASGHHTHHQSNNKLS 467
 Db 61 FLQPTSTMSQAASHSTSPSTPLSSPDSEQOSVASGHHTHHQSNNKLS 120
 QY 468 PKPGTGEVPLSLHYSTEGTSTIKLNTDEMSSASSRGISGCKSGQESFVPOS 527
 Db 121 PKPGTGEVPLSLHYSTEGTSTIKLNTDEMSSASSRGISGCKSGQESFVPOS 180
 QY 528 VQPEBGESETAPBESSSDVTKYQGVSAENPVENHINITOSDKTAKPLDNGERNDL 587
 Db 181 VQPEBGESETAPBESSSDVTKYQGVSAENPVENHINITOSDKTAKPLDNGERNDL 240
 QY 588 NLDSSCGVPEESASEKKEPETSDDQSTESATNENNTNPEPQTEATGSAHEETSTR 647
 Db 241 NLDSSCGVPEESASEKKEPETSDDQSTESATNENNTNPEPQTEATGSAHEETSTR 300
 QY 648 DSALQDITDDDDPVLIPGARYRAGPGDRSAVARIOEFFRRKREKMEELDTNTRP 707
 Db 301 DSALQDITDDDDPVLIPGARYRAGPGDRSAVARIOEFFRRKREKMEELDTNTRP 360
 QY 708 LVKAVYVGHNSRRTMKEANFWGANGFVWSSGDCGHITWDRHTAEHLMLLEADNHVNCV 767
 Db 361 LVKAVYVGHNSRRTMKEANFWGANGFVWSSGDCGHITWDRHTAEHLMLLEADNHVNCV 420
 QY 768 QPHFDPITLASSGIDYDIKIMSPLEESRIFNRKLADYITNMLLEETRNTIVPASFM 827
 Db 421 QPHFDPITLASSGIDYDIKIMSPLEESRIFNRKLADYITNMLLEETRNTIVPASFM 480
 QY 828 LKMLASLNHRADRLDGRSGSGQENENDEE 860
 Db 481 LKMLASLNHRADRLDGRSGSGQENENDEE 513

RESULT 3
 Q9P0U0 PRELIMINARY; PRT; 533 AA.
 AC Q9P0U0; 09HOM1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE PC326 protein.
 GN PC326 OR DKFZP434C2211.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADRENAL GLAND;
 RA Li Y., Shi J., Huang C., Jiang C., Ren S., Zhou J., Yu Y., Xu S.,
 Wang Y., Fu G., Chen Z., Han Z.;
 RT "A novel gene expressed in human adrenal gland."
 RL Submitted (Mar-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 52-533 FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE-21154917; PubMed-11230166;
 RA Mennemann S., Well B., Wellenreuther R., Gassenhuber J., Glassl S.,
 Mennemann S., Boecker M., Bloeker H., Bauersachs S., Blum H.,
 Lauder J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
 Mewes H.W., Oltenshaender B., Obermaier B., Tampe J., Heubner D.,
 Wambolt R., Korn B., Klein M., Poustka A.;
 RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
 Analysis of 500 Novel Complete Protein Coding Human cDNAs."
 DL EMBL; A1150234; AAF07474.1;
 DR EMBL; A1136738; CAB66672.1;
 DR InterPro: IPR000048; IO_region.
 DR InterPro: IPR001005; Myb_DNA_binding.
 DR Pfam: PF00612; IO: 1.
 DR Pfam: PF00400; WD40. 2.
 DR SMART: SM00015; IO: 1.
 DR SMART: SM00320; WD40. 2.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
 KW Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 533 AA; 58974 MW; 7A38168924477F CRC64;

Query Match 58.5%; Score 2647; DB 4; Length 533;
 Best Local Similarity 96.2%; Pred. No. 2.9e-164;
 Matches 513; Conservative 0; Mismatches 0; Indels 20; Gaps 1;

QY 348 MLSRMEFEASEVAQSNRGRSRPGTSGDSISTLPYSSPDLEVSSTAMEVDTPAQ 407
 Db 1 MLSRMEFEASEVAQSNRGRSRPGTSGDSISTLPYSSPDLEVSSTAMEVDTPAQ 60
 QY 408 FLQPTSTMSQAASHSTSPSTPLSSPDSEQOSVASGHHTHHQSNNKLS 467
 Db 61 FLQPTSTMSQAASHSTSPSTPLSSPDSEQOSVASGHHTHHQSNNKLS 120
 QY 468 DSALQDITDDDDPVLIPGARYRAGPGDRSAVARIOEFFRRKREKMEELDTNTRP 707
 Db 301 DSALQDITDDDDPVLIPGARYRAGPGDRSAVARIOEFFRRKREKMEELDTNTRP 360
 QY 708 LVKAVYVGHNSRRTMKEANFWGANGFVWSSGDCGHITWDRHTAEHLMLLEADNHVNCV 767
 Db 361 LVKAVYVGHNSRRTMKEANFWGANGFVWSSGDCGHITWDRHTAEHLMLLEADNHVNCV 420
 QY 768 QPHFDPITLASSGIDYDIKIMSPLEESRIFNRKLADYITNMLLEETRNTIVPASFM 827
 Db 421 QPHFDPITLASSGIDYDIKIMSPLEESRIFNRKLADYITNMLLEETRNTIVPASFM 480
 QY 828 LKMLASLNHRADRLDGRSGSGQENENDEE 860
 Db 481 LKMLASLNHRADRLDGRSGSGQENENDEE 513

OY 748 RHTAEHMLLEADNHVNCIOPHPDPIILASSGIDYDIKIMSPLEERSIFNRKLADEVIT 807
 DB 421 RHTAEHMLLEADNHVNCIOPHPDPIILASSGIDYDIKIMSPLEERSIFNRKLADEVIT 480
 OY 808 RNEIMLEETRNITVPASFMRLASLNIHRAIRLGRSGSGSGGONENDEE 860
 DB 481 RNEIMLEETRNITVPASFMRLASLNIHRAIRLGRSGSGSGGONENDEE 533

RESULT 4

OY 090UD5 PRELIMINARY: PRT: 485 AA.
 AC 090UD5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE D703H14.1 protein (Fragment).
 GN D703H14.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Grafting D.;
 CC Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL; AL031287; CAB52430.1;
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 5.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 DR Repeat; WD repeat.
 KM NON_TER
 FT SEQUENCE 485 AA; 54745 MM; F47AF430351AE219 CRC64;
 SQ

Query Match

Best Local Similarity 49.9%; Score 2258; DB 4; Length 485;
 99.8%; Pred. No. 5.3e-139;
 Matches 426; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 RREFIORLKEATLVNHDGCVNTICMNDTEYILSGSDTKIYINPYSRKVLTTIRSGH 93
 DB 1 RREFIORLKEATLVNHDGCVNTICMNDTEYILSGSDTKIYINPYSRKVLTTIRSGH 60
 OY 94 RANIFSAKFLPCTNDKQIVSCSGDGYIFTYNEODAEINROCQFCHYTYEIMTVND 153
 DB 61 RANIFSAKFLPCTNDKQIVSCSGDGYIFTYNEODAEINROCQFCHYTYEIMTVND 120
 OY 154 PYFPLSCGDEGVYRWFDTRIKTSCTRECKDDILNCRRAATVAICPPITYLAVGSD 213
 DB 121 PYFPLSCGDEGVYRWFDTRIKTSCTRECKDDILNCRRAATVAICPPITYLAVGSD 180
 OY 214 SSYRIIDRRLGTRATGNVAGRTGTGMVAFIPSHLNNSCWTSLCYSEDOELIVYS 273
 DB 181 SSYRIIDRRLGTRATGNVAGRTGTGMVAFIPSHLNNSCWTSLCYSEDOELIVYS 240
 OY 274 SDYIYLFDPKDTARELKTPSAEERREELROPVYKRLRLGWDSDTPRAPRESERERG 333
 DB 241 SDYIYLFDPKDTARELKTPSAEERREELROPVYKRLRLGWDSDTPRAPRESERERG 300
 OY 334 EOSPNSVLMORMSDMLSRWFEASVYAQSNRGRSRPRGTSQSDISTLPYVSSPDLE 393
 DB 301 EOSPNSVLMORMSDMLSRWFEASVYAQSNRGRSRPRGTSQSDISTLPYVSSPDLE 360
 OY 394 VSEIYAMVDTFPAQFLOPSTSSMSAOAHSTSPSPHSTPLSSPDSEBOROVASGH 453
 DB 361 VSEIYAMVDTFPAQFLOPSTSSMSAOAHSTSPSPHSTPLSSPDSEBOROVASGH 420
 OY 454 HTHHOSD 460
 DB 421 HTHHOSD 427

RESULT 5

OY 09LEU6 PRELIMINARY: PRT: 754 AA.
 AC 09LEU6;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical 83.5 kDa protein.
 GN T30N20.210.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL; AL365234; CAB96849.1;
 DR InterPro; IPR000345; Cytic_heme_bind.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 6.
 DR Prodom; PD000018; WD40; 1.
 DR SMART; SM00320; WD40; 6.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
 DR PROSITE; PS50082; WD_REPEATS_2; 2.
 DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
 KM Hypothetical protein; Repeat; WD repeat.
 FT SEQUENCE 754 AA; 83526 MM; C63FC6E7608B2101 CRC64;
 SQ

Query Match

Best Local Similarity 13.0%; Score 590.5; DB 10; Length 754;
 24.6%; Pred. No. 3.4e-30;
 Matches 207; Conservative 11; Mismatches 279; Indels 245; Gaps 28;

OY 22 EDPGRLRSRYLGRREFIORLKEATLVNHDGCVNTICMNDTEYILSGSDTKIYINP 80
 DB 20 EDPGHEVDQRKQFISLVRSLSGDELEHGQCNALAMNSNLSISGDDLRINIMNY 79
 OY 81 YSRKVLTTIRSGHRAIPSAFLPCTNDKQIVSCSGGVY--TYTN-----VQDAETNR 133
 DB 80 SSRKLHSDIDGHANIFCTFVPEISDELVSAGAEVRLFTSLSGRADDNAIIP 139
 OY 134 QCFYCHGTYEIMTVNDYFTLSCGDEGVYRWFDTRIKTSCTRECKDDILN 188
 DB 140 SALVQCHRRYKKAIVAEVGNPNVYMSASDEDTLQOHDFRESTSCPAGTNAQECR -SVLL 198
 OY 189 NCRRAATVAICPP-----IPYLAAGCSGSDSVRIYDRMLGTRATG----- 230
 DB 199 DLRSAGKALADPRQQLSLKSCDISATRPMLILVYGSDAFARLYDRMRPLPLSSKR 258
 OY 231 -----NAGSGTGMVAFIPSHLNK-----SCVTSICYSEDOELIVYSYIYLF 281
 DB 259 PPPPCVNY-----FCPMHLSERGRTNLHITHTFSPNGBEVLLISGGEVILMN 307
 OY 282 PDDOTARELKTPSAEERREELROPVYKRLRLGWDSDTPRAPRESERERGOSPNVSL 341
 DB 308 VNNGTGIMQYTPGDVNLFSF-----SNNLHDVESPQVYS- 342
 OY 342 MORMSDMLSRWFEASVYAQSNRGRSRPRGTSQSDISTLPYVSSPDLE-EVSEIAME 400
 DB 343 -----TPQNGFHRSS--SNAATYKCTCLELVEIAKLSLE 373
 OY 401 VDT-----PAQFLOPSTSSMSAOAHSTSP-----TESPHSTPLSSPDSE 443
 DB 374 EGTDFYALAEANEVLAHNSDIESALRHECICTRAALLKRNKNDH-----NAVROCH 429

```

QY 444 QKQSVASCHHTHHOSDNNNEKISPKPGCEPVLSTHYEGTTSTIKLNFTEWSSIA 503
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 NNRRIDASSFKAHYVSEALQOL-----GKCKEALDFATAQHMPDADIAKVESIK 483
QY 504 SSSRIGSHCKSGEGESFVPOSSVOPPGDSSTK-----APESEEDVTKYQ 551
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 RLQQAAG-----AKNET-----GAGTTRVLSLIDILYREANSDDSHMSRSE 528
QY 552 -EGVSAENVENHNINTOSDKFTAKPLDNGSERNDLNDRCGVPEESASSEKAKEPET 610
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 RSDSDVEDELIDTQSLSD-EGRDTDSN-MRGSINL-----RIHVGDDKPEE-NT 579
QY 611 SQTSTESTNNNTNPEQFQTEATGPAHESTRSALDPTDSDDDPVLIPGARV 670
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 580 VDNASSGTSASSQN-----DRTSYQPEGALD-----605
QY 671 AGPDGRSAVANIOEFFRRKREKMEELDTLIRPLVKVYKGRNRTMIKEANFNG 730
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 -----MKRRYVGHCVNGTIDIKQASPLG 627
QY 731 --ANFVMSGDCGHIFIMDRHTAEHLMLLEADNHYVNCLOPHFPDPIILASSGIDYDIK 788
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 628 QGGEYIASGSDGRFWFIWEKQGRIMKVLVGDESVLNCIQCHPDPVSVAATSGIDNTIK 687
QY 789 SP 790
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 688 SP 689

RESULT 6
ID 094B03 PRELIMINARY: PRT: 757 AA.
AC 094B03:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 83.8 kDa protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koeseema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shino P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full length cDNA of gene T30N20.210/At5g10940 (GI:8979728).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: AY039964; AAK64141.1;
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001005; Myb_DNA_binding.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 6.
DR ProDom: PD000018; WD40; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS50082; WD_REPEATS_2; 2.
DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 757 AA; 83830 MW; 12BE45AAC114056D CRC64;
Query Match 12.8%; Score 580; DB 10; Length 757;
Best Local Similarity 24.5%; Pred. No. 1.7e-29;
Matches 207; Conservative 111; Mismatches 279; Indels 248; Gaps 29;

```

```

QY 22 EDS-RLRSRYLGRREIORTKLEATLVNDGCVNTICWMDTGYILLSGDDTKLVISNP 80
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 QDPSHEVDQMOQSHSILVRLSQEQLGHOQGVNMAIAMSNSLLISGSDDIRINIMNY 79
QY 81 YSRKVLITTINGHRANIFSAKFLPCTNDKQIVSCSGGV--PYTN-----VEQDAETNR 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 SSRKLHSLIDTGHANIFCKFYFETSDLVSGAGDAEFLRNLTSLSRAEDNAIIP 139
QY 134 QCFPTHGCTTYEIMVYNDPYFLSCGEDGVRFMDTRIKTSC-----TKEDCKDILI 188
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 SALYQCHTRKRVKLAVEPNPNVWVWASSEDGTLRHDFRESTCPAGTAHQCR-SVLL 198
QY 189 NCRRAATSVACPP-----IPYLAAGCSQSVRYTDRMLGTATG-----230
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 DLRSAGKRALADPKPTLSKCDISATRPHLLVGSQDPFALYDRMLPIIASSRKRM 258
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 -----NYAGRTGGVAVRPIPSHLNKK--SCRYTSLCYSEDQELIVSYSDYITLFD 281
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 PPPPCVNY-----FCPMHLSEGRNTLHLHYTFSPGGEVLLSYGEHYITAN 307
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 PKD---DTARELTPSAEERRELRQPYKRLRLRGDWSDTGRAPRESERERDQSPN 338
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 VANGICSTIMQTTPEQVNLFSF-----SNNLHDVESPPQ 343
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 VSLMORMDMLSRWFEASEVAQSNRGSRGRPRGTSQSDISTLPYVSPDL-EVSET 397
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 344 VS-----TTPQNFHRS--SNATVAKCTELVETAKW 373
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 398 AMEVDI-----PAQFLQPTSSSTMSQAHOSTSP-----TESPSTPLSSP 440
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 SLEEGTVFYALEANVEIDAHNSDIESALRHCLOTRAALLIKRKMDAH---MAVR 429
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 DSEOROSVASGHHTFHOSDNNNEKISPKPGCEPVLSTHYEGTTSTIKLNFTEWSS 500
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 DCHNRRIDASSFKAHYVSEALQOL-----GKCKEALDFATAQHMPDADIAKVE 483
QY 501 STASSRIGSHCKSGEGESFVPOSSVOPPGDSSTK-----APESEEDVTKYQ 548
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 SLKRLQAAG-----AKNET-----GAGTTRVLSLIDILYREANSDDSHMS 528
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 549 KYQ-EGVSAENVENHNINTOSDKFTAKPLDNGSERNDLNDRCGVPEESASSEKAKE 607
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 RSDSDVEDELIDTQSLSD-EGRDTDSN-MRGSINL-----RIHVGDDKPEE 580
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 608 PETSDQSTESATNNNTNPEQFQTEATGPAHESTRSALDPTDSDDDPVLIPGA 667
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 581 -NTVDMASSGTSASSQN-----DRTSYQPEGALD-----608
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 668 RYRAGPDGRSAVANIOEFFRRKREKMEELDTLIRPLVKVYKGRNRTMIKEAN 727
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 -----MKRRYVGHCVNGTIDIKQAS 627
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 728 FWG--ANFVMSGDCGHIFIMDRHTAEHLMLLEADNHYVNCLOPHFPDPIILASSGID 785
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 628 FLGQGEYIASGSDGRFWFIWEKQGRIMKVLVGDESVLNCIQCHPDPVSVAATSGIDNT 687
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 786 KIWSP 790
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 688 KIWSP 692

RESULT 7
ID 094527 PRELIMINARY: PRT: 809 AA.
AC 094527:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative guanine nucleotide binding protein beta subunit-like.
GN SPBC609.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;

```

OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AL035226; CAA22832.1;
 DR InterPro: IPR001230; Ptenyl_site.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.5.
 DR Prodom: PD000018; WD40.1.
 DR SMART: SM00320; WD40.5.
 DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
 KW Repeat; WD repeat.
 SQ SEQUENCE 809 AA; 92498 MW; 4877443546BD057E CRC64;

Query Match 12.1%; Score 547.5; DB: 3; Length 809;
 Best Local Similarity 24.8%; Pred. No. 2.4e-27;
 Matches 236; Conservative 124; Mismatches 308; Indels 285; Gaps 41;

27 LRSRYGREGREIOR-----LKEATLVHDCGVTICWNTGTEYILSGSD 71
 6 LLSRQLDYDMQKRSRITGNSWTWLTIDQLKELTGTCVNTLMDADDFLLSGSD 65
 72 DTKLV---ISNPYSKRYLTTRSGHRANIFSAKFLPCTNDKQIVSCSGDGVITYNVED 128
 66 DTRLIYVDFNEKPRHL--ISTGHQVNIKSAFVYSNNRQILSASGDKLRLDSDS 123
 129 -----AETNQCQFCHGTYTIVMVPNDPVTEFLSCGEGYRWMDTKLTKSCTE 180
 124 KEGMDHMETQTRC--WSCALDSKNIVPCDNC--HTELVCSBEGDTRKQYDIPEHYCND 181
 181 -DCKDILLIN---CRRATSVAILCPPIPYLLAVGSCDSVRIRYRRLMGTTRA---TGNV 232
 182 LDC-PSLIVVYNVYRLNLTITWSPNPYFALIGTHPYAFLYDRMVKKSPFDMWTMT 240
 233 AGRGTTGMVAKFLPSSLINKSC-----RVSLCT--SEDOGLIVYSSSYIYFDPK 263
 241 SPKDCRCVAKFSP---DSCNSQGLIDRYITCCGFSANPELLIWNMSDVIYLFVHV 296
 284 DDTARELKTSPASERREBELKOPVVKRLRLGWDSDTGPAPRPEREREDGEOSPNVSLMO 343
 297 EDKS---YPTFKIEDSKKPKPSL-----LQTPPLKRRK-----330
 344 RMSDMLSRWFEASEVAQSNRGRGRSRPRGGTQSODISTLPYVSSPDLEVEETAMEVDT 403
 331 ---YSPWY-----KNMF-----AST 343
 404 PAQDFLOPSTSSMTSAQHSSTPSPHSTPLSSPDSPQSVASGHHTHGSDNNK 463
 344 PA-----SRVSNPPTAAQPRK-----HTTYQYEMT 370
 464 EKL--SPKPGTGEVLSLHSTEGTTSTIK--LNFTEWSSIASSS-----506
 371 EKFTTNGGLVGSIVSGRLSHFSRSIQYKDAIYFLENNTYIPDSNGLNHSIRVSALRY 430
 507 -RGIGSHCKSEGOESFVPOSSVQPPRGDEETAPRESSADYKQYEGV-----A 556
 431 WRACVSLTALMDVTSLDEPTITIOAGWG-----WLYDFNMWVTRYLGLISDHVALQMSPP 485
 557 ENPVENHINTQSDKFTAKPDSNGE-----RNDLN-----LDRSGVPE 597
 486 TNVAKQNVLCDDPE-PSRVLFSRPSMIRAFRIDNDLSVRFRFVHKVLRGCLLLI 544
 598 ESASSEKAKEP---ETSDQSTESATNEN--TNEPQFQTEANGPS-----AH 641
 545 SSDIYWEQFQWDSSTSDVTSISQRLDDENGFLTLLEPPVNYEMNEVSSGENTIVSMYTG 604

OY 642 EETSTRSALQD-----TDSDDDPVLIPGARVYRAGDRRSARVARIQEFF 687
 DB 605 SLDNDDDDQYODEESYASDDEDDDESDDEGPITLS-----642
 OY 688 RRRKRRKEMEDLTNIRRLPYKMYKGRNSRTMKIKANFWCAN--FYMSGDCIHFI 745
 DB 643 LRMKRAVPPNVVNT--VKSXY-GHCNVS-ITKNVYQNDQYVMSGSDDRFET 697
 OY 746 WDRHTARHMLLEADNVVNCLOPHPPDPITLASSGIDYDIKWS-----PLESRIFNKR 800
 DB 698 WDKLNASTLITIHGDSAAVAVIECHPCPLTAVSGIDSYTKIFNTENTPPSCGRNHTSN 757
 OY 801 LADEVITRNEIMLEETRNITVPASFMLRLMLASLNHIRADRLGRSDSGSGOE 853
 DB 758 -SYKTIATFNEKNRQGSGRDSYI--TSRMLSHLAVRAHL-----DDGGRHE 799
 RESULT 8
 ID O9V8C7 PRELIMINARY; PRT; 628 AA.
 AC O9V8C7;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
 DE CG5124 protein.
 GN CG5124.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.A., Bouck J., Brokstein P., Brotler P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dushin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasteo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Peltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitzkas R., Tector C., Turner R., Venter L., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).


```

Db 290 NMGYG----- 294
QY 423 SSSPESPSPHSTPLSSPDEQROSVASGHHHTHOSDNNNEKLSPKPGTGPVLSLHYS 482
Db 295 -SSPV----- 298
QY 483 TRGTTSTIKLFTDMSIASSSRGSHCKSEGOESFVPOSSVOPPEGSETRAPEE 542
Db 299 ----- 298
QY 543 SSEDVTKYQGVSAENPVENHINTTOSDKFYAKPLDSSNGERNDLNDSGCVPESSAS 602
Db 299 ----- 302
QY 603 EKAKPETSDQSTESATNENNTNPEPOTATGP5AHEETSTDSALQDTDDSDDDV 662
Db 303 EKLQEME----- 312
QY 663 LIPGARYRAGPGDRRSARVARIOEFFRRRKRKEMEELDTLNRPLVKVYKGRNSRTM 722
Db 313 ----- 322
QY 723 IREANTWGAN--FVMSGDCGHFTFDRTAHEHMLLEADNHYVNCLOPHPEPDTIASSG 780
Db 323 VGVNPFEGNDEYVTSGDSOGHIFIKKKGKGLVRAMVGDRRVNOLSESHPIPLASCG 382
QY 781 IDYDIKIMPLEBSRIENRKLAEVITRNELMLEETRNITVPAFPM--LRMLASLHNR 838
Db 383 IESVAKLTPMSNDVLSLEPKIDKVELNRYGVEDOSRTVLTDPVIMHVLRLQROTSAF 442
QY 839 ADRLBGRSESGQENE-----NEDE 859
Db 443 TERRVYSTDIGSDGDEGDAFIAVLANDDE 472

RESULT 10
Q12839 PRELIMINARY; PRT; 597 AA.
AC Q12839;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE H326 protein.
GN H326.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MYELOMA;
RA Bergsagel P.L.; Kuehl W.;
RT "H326 is a human gene homologous to murine PC326 that is ubiquitously
RT expressed."
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL: U06631; AAA16607.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00320; GPROTEINBRPT.
DR PRODOM: PD000018; WD40; 1.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
DR REPEAT: WD repeat.
SQ SEQUENCE 597 AA; 66384 MW; C0C696AD603645B9 CRC64;

```

Query Match 9.8%; Score 443.5; DB 4; Length 597;
 Best Local Similarity 18.1%; Pred. No. 9.3e-21;
 Matches 161; Conservative 83; Mismatches 173; Indels 471; Gaps 17;

```

QY 2 SRGSGYPHLLMD----VRKSLGLEDEPSRLRSYLGRRREFIQRILKEATLNVHDCGVNTI 57
Db 142 SETSALPRPRWQALPALREBEL--SSARFVYACAGARVYVHGFRHQHLEGHTGCNTL 199
QY 58 CMMDTGEYILISGSDDTIKLVISNPFYSKRVLTITSSGRANIFSKFLPTCDKQIVSCSD 117
Db 200 HFNQRTWLASGSDDLKVVYMDVWRROPVLDPEFSGHSNFFQAKFLPNSGSDSTLANCARD 259
QY 118 GVIFYTNVEOD--AETNROCQFTCHYGTVEITFVNDPYTFLSGEDGTVMFPDR-- 172
Db 260 GQYRAVELSATQCCNKTKRAQ--HKGASHKALALPDSCTILSAGEDAVYFTIDRLD 316
QY 173 --IKTSCYKEDCKDILNCRRAATSVACPIPYLLAVGCSDSVRIYDRMLGTRAT 229
Db 317 RPASKLVVTKEREK-----KVGLYTIYVNPANTHQAFAVGDRQFRIYDQKIDENE- 368
QY 230 GNTAGKGTGMVARFTPSHLNKSCL--VTSLEYSGGGLVYSYSDYLYLDPKDDTA 287
Db 369 -----NNGVLKFCPHHLVNSKANTICLVYSHDGETELASYNDEDIYLRN----- 415
QY 288 RELKTPSAEERRELRQPVKRLRLRGDMSDTGPAPARESERERDQSPNVSLQRMSD 347
Db 416 ----- 415
QY 348 MLSRWEASEVAQSNRGRSRPRRGTSQSDISTLPYVSSDLVEYSETAMEVDPAEQ 407
Db 416 ----- 415
QY 408 FLQPTSTSTMSAQHSTSPTESPHSTPLSSPDEQROSVASGHHHTHOSDNNNEKLS 467
Db 416 ----- 415
QY 468 PKPGTEPVLSLHYSTEGTTSTIKLFTDMSIASSSRGSHCKSEGOESFVPOSS 527
Db 416 ----- 415
QY 528 VQPEGDSFKAPESESEDVTKYQEGVSAENPVENHINTTOSDKFYAKPLDSSNGERNDL 587
Db 416 ----- 415
QY 588 NLDRCGVPEESASSEKAKPETSDQSTESATNENNTNPEPOTATGP5AHEETSTR 647
Db 416 -----SSHSD-- 420
QY 648 DSALQDTDDSDDPVILPGARYRAGPGDRRSARVARIOEFFRRRKRKEMEELDTLNR 707
Db 421 -----GQY-----YKR- 427
QY 708 LVKMYTKGHRNSRTMIKANFWG--ANFVMSGDCGHFTFDRTAHEHMLLEAD-NHYV 764
Db 428 -----YKGHRNNAF-VKGVNFYGPKEFVYSGSDCGHIFLMEKSSCOIQFMGEGDKGVV 481
QY 765 NCLQPHPEPDIILASGDVYDIKIMSPLEBSRIENRKLAEVITRNELMLEETRNITVPA 824
Db 482 NCLPPEHPLVLAFTSGDHDVYKIMAPYAEASTELTGKVDYKKNKRERDSDSLHQDTLFD 541
QY 825 SFMLRLMASLHNRADR-----LEGDRSESGQENENDEE 860
Db 542 SHMLWFL--MHLRLQRHHRHMRREPVGATDADSDSPSSDSDSEEE 587

RESULT 11
Q01078 PRELIMINARY; PRT; 747 AA.
AC Q01078;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Protein PC326.
GN PET2 OR PROTEIN PC326.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI_TaxID-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=PLASMACYTOMA;
 RX MEDLINE=93026383; PubMed=1408147;
 RA Eckhardt L.A., Laakov R., Kuehl W.M., Bergsagel P.L., Timblin C.R.;
 RT "Sequence and expression of a murine cDNA encoding PC326, a novel gene
 expressed in plasmacytomas but not normal plasma cells."
 RL Oncogene 7:2059-2064(1992).
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: M05564; AAA39895.1; -.
 DR MGD: MGI:101758; Pet2.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 6.
 DR ProDom: PD000018; WD40; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 DR Repeat: WD repeat.
 KW SEQUENCE 747 AA; 83193 MW; 48E40B6378BD78A CRC64;
 SQ
 Query Match 9.5%; Score 429.5; DB 11; Length 747;
 Best local similarity 18.7%; Pred. No. 1e-19;
 Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;
 OY 8 PHILMVDY---RKRSLGLEDESRRLRSRYLGRREFIQRKLEATLVNHDGCVNTICMNDTG 63
 DB PRPMNVYLSALRDRLQ--SSGRFYEEACGARLFVGRFSLHVEFGHSCVMTVHFNQHG 365
 OY 64 EYLLSGSDTLVYNSPFSKRVLTIRSGHRANITSAPKLPCTNDKQIYSCSGDGYIFT 123
 DB 366 TLASGSDTLVYVMDWIKKRSVLFNDSGKNNILQAKFLPCNDLILMCGDGVRYA 425
 OY 124 NVEQDAETNRCQCFCHGYEIMTVPNDEPTFLSCGEGVWMEVRIKTSCTEKEDK 183
 DB 426 QLSAVAGHMKRLVKHGGASHRLGLEPDSRPFRLTSGEDAVENIDLQAHASK----- 481
 OY 184 DDILI---NCRRAATVAICPPIDYILAVGCGSSSVRIYDRMAGTRATGNTAGCTTG 239
 DB 482 --LLVIKGDKRKVGLYFVFNPNANYOFAPVGGDQFMRIYDQKIDENV-----NNG 531
 OY 240 MVAFPIPSHL--NNKSCVTSICVSGEQLVYSYSSDIYLFDPKDTARELKTPSAE 297
 DB 532 VLKFCFPHLLSSDPPIAHITSLMYSTDTGELLASTNDEDIYF----- 574
 OY 298 RREELROPVYKRLRLKGDWSDTGPRAPRESEREERDEQSPNVSLAQMSDMLSRMEAS 357
 DB 575 ----- 574
 OY 358 EVAQSNRGRGSRPRGTSQSDISTLPYVSSPDLVSETAMEVTPAQFLQPSSTM 417
 DB 575 ----- 574
 OY 418 SAQAHTSSPSPHSTPLSSPDSERQSVASGHTTHGSDNNNEKLSPKPGTGPVL 477
 DB 575 ----- 574
 OY 478 SLVSTEGTTSTIKLFTDEMWSIASSSRGIGSHCKSEGGQESFVQSSVOPPEGDSET 537
 DB 575 ----- 574
 OY 538 KAPESSSEDVYKQGVASNEVENHINTQSDKFTAKPLDLSNGERINDLRSCGVPE 597
 DB 575 ----- 574
 OY 598 ESASSEKAKEPETSDQSTESATNENNTNPEQFQTEATGSAHETSTRDALSQDTPDS 657
 DB 575 -----NSDS 579
 OY 658 DDDPVLIPGARYRAGGDRRAVARIQEFFFFRRRKRKEMEELDTLNRPLVMKYKGR 717
 DB 580 D-----CAQY-----AKR-----YKGRH 592

OY 718 NSRTIKENAFNG--ANFVSSGSDCGHIFTMDRTAHEMLLEAD-NHVNCLQHPEDP 774
 DB 593 NNST-VKGVYFYPGRSEFVSGSDGHIETWESSCIVQFLEDEGCTINCIDSHPLP 651
 OY 775 ILASGIDYDIKTPSPLEESRIFNRKLA--DEVTRNLMEE--TRNTIVPAS---FM 827
 DB 652 VLASGIDHEVYKTPSPFAEP---SKLAGLKNVYKIKRKNKNDNTLRITSLFNNSMLCFL 708
 OY 828 LRMILASLNH-----IRADRLGDRSGSGGGENENEDE 859
 DB 709 MSHVYQSNYGRSMWRGIRINAGGDFSDSSSSSEETNOE 746
 RESULT 12
 Q9T001 PRELIMINARY; PRT; 493 AA.
 ID O9T001;
 AC O9T001;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 54.9 kDa protein.
 GN T12J5.10 OR AT4G35140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Hohnsels J., Mewes H.W.,
 Mayer K.F.X., Schueller C.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Purnelle B., Boutry M., Goffeau A., Mewes H.W., Lemcke K.,
 Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AL035522; CAB36721.1; -.
 DR EMBL: AL161586; CAB80231.1; -.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 6.
 DR ProDom: PD000018; WD40; 1.
 DR SMART: SM00320; WD40; 3.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
 DR Hypothetical protein; Repeat; WD repeat.
 KW SEQUENCE 493 AA; 54937 MW; CF8F5C207CB48BAA CRC64;
 SQ
 Query Match 9.0%; Score 405.5; DB 10; Length 493;
 Best local similarity 18.9%; Pred. No. 2.1e-18;
 Matches 162; Conservative 74; Mismatches 151; Indels 471; Gaps 20;
 OY 9 HLMVYRRKSLGLEDEPSRLRSYLRGRFIOQLKLEATLVNHDGCVNTICMNDTGEYIS 68
 DB 17 HPYVDVFRREVYGGISRNFSNRSFASENVLRLEIKKLEKIKKGCYNTVSFNAEGDYLLS 76
 OY 69 GSDTKLIVISNYSRKVLTITISGRANIFSAKFLPCTNDKQIYSCSGGCVFYFTVEQD 128
 DB 77 GSDRVRVLMWQIGNVKLSFHSGHANNVFOAKFMFSDRIVTCADAGMF--DLRTE 133

```

QY 129 AETNRQCOPTCHYGTYTEIMTVNPDPYTLFSCGSDTVAWFDTRIKTCTCKEDCKDDILLI 188
DB 134 APTC-----LFTC-----RSVDPRRRN-----MDAIOI 156
QY 189 NCRRAATSAICPPPIPYLAAGCSDSVRIYDRMLGTATNGYAGRTGTVYA---RI 245
DB 157 N-----AIAIDPRSNLEFAVGMEYALYDIR-----RFOGEGINGFTTAAADHFC 202
QY 246 PSHL-NKSCRAVTLSCYSEDGGEILVSYSDYTYLDPKDDYAREIKTPSAEERREELQ 304
DB 203 PPHLIGNEVDVIGITGLAFSEQ--SELLVSYNDEIFYLFTP----- 239
QY 305 PPVKRLRLRGWSDTGPAPAREPSEERROCEOSPNSILMQRMDSMLSRWFEASEVAQSNR 364
DB 240 ----- 239
QY 365 GGRSRPRGATSQSDISTLPVPSPPDLEVSETAMEVDTPAQFLQPTSTSTMSAQAHS 424
DB 240 GMGLG----- 244
QY 425 SSPTESPSTPLSSPDSEORQSVASGHHHTHQSNNNEKLSPKPGTGEVPLSLHYSTE 484
DB 245 -----SNIPSSPIS----- 254
QY 485 GTTSTIKLNTDEMSSIASSSRGICSHCKSGQESFPVQSSVQPPREGDSETKAPRESS 544
DB 255 -----KSPVSKS 261
QY 545 EDVTKYQGVSAENFENHINITQSKFTAKPLDSNGERNDLNDRCGVPESASSEK 604
DB 262 -----ESSSPK 268
QY 605 AKPEPISQPTESATNENNTNPEQFQTEATGSAHEETSTRSALODTDDSDDPVLI 664
DB 269 DE-----NHS----- 274
QY 665 PGARYRAGPDRRSAAVARIOEFFRRKREKMEELDITLIRRPVLYKMYKGRNRGRTMK 724
DB 275 -----VSLYKKGHKNCET--VK 289
QY 725 EANTWG--ANFVMSGSCGHIFFINDRTAEHLMLLEADNHVNCLOPHFDPITLASSGID 782
DB 290 GVNFPGPSEVYVSGSDGRIFIRKKKGELIRVMEADRHVNCLEPHPIVPLASSGIE 349
QY 783 YDIKMSPLEESRIENRKLADDEVITRNMLMEFRNTTIVASFTMLASLN-HIRADR 841
DB 350 SDIKVMT-----SKAERATLPE-----NIELPSRFRIPVLSLSEHYDDE 392
QY 842 LEGDRSESGQENENEDE 859
DB 393 LFGNGMDIGIDNGEDE 410

```

RESULT 13

Q9SZM8 PRELIMINARY; PRT; 426 AA.

AC Q9SZM8; 01-MAY-2000 (TReMBUrel. 13, Created)

DT 01-MAY-2000 (TReMBUrel. 13, Last sequence update)

DT 01-MAR-2002 (TReMBUrel. 20, Last annotation update)

DE Hypothetical 47.5 kDa protein.

GN F20M13.40 OR A74G38480.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Medler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W., Mayer K.F.X., Scheller C.;

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

[2]

```

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Medler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: CONTAINS 7 WD REPEATS (WRP-ASP DOMAINS).
DR EMBL: AL035540; CAB37504.1;
DR EMBL: AL161593; CAB80512.1;
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR ProDom: PD000018; WD40; 1.
DR SMART: SM00320; WD40; 5.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 426 AA; 47451 MW; 0A6050076B8ECECA CRC64;

```

Query Match 8.5%; Score 384.5; DB 10; Length 426;
Best local similarity 17.6%; Pred. No. 4,1e-17;
Matches 137; Conservative 63; Mismatches 116; Indels 461; Gaps 14;

```

QY 32 LGRREFIORKLEALNHDGCVNTICMNDGELYLLSGSDTKLYISNPKVLTTRS 91
DB 3 LOKDRLRLGDKRLDKHKGCVNTVSFNADDLILSGSDRQVILIMQZASVRLSEDS 62
QY 92 GHRANIFSAKLEPCTNDKQIVSCSGDGYLYTNEQDAETNNQCOFTCHYGTYTEIMTV 151
DB 63 GHENNIFOAKMPEFSDDBDTIYTSADKQYRSKILESQVETSL-LGKHQGVHKLAVEP 121
QY 152 NDPTFLSCGEGYVR-----WFDTRIKTCTCKEDCKDDILLINCRRA-----AT 195
DB 122 GSPSEFYTCGDBGAVKHCIFHEVFLSTRVATN-----LFTCKEAFNLVYLH 171
QY 196 SVAICPPPIPYLAAGCSDSVRIYDRMLGTATNGYAGRTGTVANAFISHLNKSCR 255
DB 172 AIAVDRNPGILAAVAGMEYARVDIR--STRSEDDHG----- 208
QY 256 VTSLCYSEDGGEILVSYSDYTYLDPKDDYAREIKTPSAEERREELQPPVKRLRLGSD 315
DB 209 ITGLAFS-DQSELLASYSDEFILFTP----- 234
QY 316 WSDTGPAPAREPSEERROCEOSPNSILMQRMDSMLSRWFEASEVAQSNRGRSRPRGRT 375
DB 235 -----DM-----GLG----- 239
QY 376 SQSDISTLPVPSPPDLEVSETAMEVDTPAQFLQPTSTSTMSAQAHSSTSPSPHSTP 435
DB 240 -----PTPPSSSTK 248
QY 436 LLSSPDSEORQSVASGHHHTHQSNNNEKLSPKPGTGEVPLSLHYSTEGTTSTIKLNF 495
DB 249 ----- 248
QY 496 TDWMSIASSSRGICSHCKSGQESFPVQSSVQPPREGDSETKAPRESSDPTVYQEGVS 555
DB 249 -----TEERTPTP----- 256
QY 556 AENFENHINITQSKFTAKPLDSNGERNDLNDRCGVPESASSEKAKEPETSQTS 615
DB 257 ----- 256
QY 616 TESATNENNTNPEQFQTEATGSAHEETSTRSALODTDDSDDPVLIIGARYRAGPD 675
DB 257 ----- 256
QY 676 RRSAAVARIOEFFRRKREKMEELDITLIRRPVLYKMYKGRNRGRTMK 733

```


Db 257 -----VKHEHTNRET-VKGVNFEFGPKCEY 279

QY 734 VMGSDGCHIFIMDRHTAEHLMLADNHNVCLOPHFDPILASSGIDYDIKIMSP 790

Db 280 VVSGSDGCRIFIMRKKGELLNAMEADRHYVNCIESHPHMLMCGSGIDITDIKIMTP 336

RESULT 14

Q8SX92 PRELIMINARY; PRT; 743 AA.

AC Q8SX92; 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE G828796P.

GN CG8001.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY.

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,

RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Ceinliker S.;

RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY094738; AAA1091.1;

SQ SEQUENCE 743 AA; 82751 MW; 96B713D71FC95EFF CRC64;

Query Match 8.2%; Score 373; DB 5; Length 743;

Best Local Similarity 17.3%; Pred. NO. 4.8e-16;

Matches 14; Conservative 73; Mismatches 163; Indels 466; Gaps 14;

QY 31 YLGRREFIOLKLEATLHVDGCVWTCIMNGEYILSGSDDTKLVIISNPYSHRYLTTR 90

Db 301 YVSGROYVEQWTLSSLVHVGCVSNLNFNRAGDLICSSDULTIYVMWMAKOLHRRF 360

QY 91 SGHANIFSAFLPCTNDKQIYSCSGDVIYFTINVEDAETNRQOCFTCHYGTIEMV 150

Db 361 SGHMMNIFOTKFDISAGCLDIYSSSRDQVRSVIPPSCGVKIPRLTYHSEVHKIIV 420

QY 151 PNDYTFELSGEDGVHMFEDPFIKTSCTKECKDKDILINC-----RRATSVAI 201

Db 421 PHSHEMLMAGEDAAVAFEDLRASNAAT-----TMMRCYTNDEMERGRVLFSTAH 473

QY 202 PIPYLLAVGSDSVRIYDRMLGTRATGVNAGRGTTGVARFIPSHLNKSCRVTS 261

Db 474 YAPECVSG-SDDILRYVDKRL-AKAIHOMAPRNL-----LEAQITQICAVY 520

QY 262 SEDGEIIVSYSSYIYLPDKDDTARELKTSAERREELNQPYPYKRLRLRDMSD 321

Db 521 NMSGSEIIASYDAGIYLFDSRNYN-----RQDY----- 549

QY 322 RARPERERDEQSPNVSLMQRMSDMLSRWFEASEVAQSNRGRSRPRGTSQSDIS 381

Db 550 ----- 549

QY 382 TLPTVPSSPDLEVESEAMEVDPAPAOFLQPSYSSMSQAQAHSTSPESHSPPLSSPD 441

Db 550 ----- 549

QY 442 SEQROSVASGHHHTHQSNNNEKLSPKRGTPGVPVLSLHSTGTTTSTIKLNFDD 501

Db 550 ----- 549

QY 502 IASSRGIGSHCKSEGOESEFVPOSSVOPPEGDESETKAPESSESDVTKYOGVSAEN 561

Db 550 -----LHC----- 552

QY 562 NHINIQSDKFTAKPLDSNSGERNDLIDRSCGVPEESASEKAKEPETSOTSTESATN 621

Db 553 ----- 552

QY 622 ENNTNPPOFOTATGSAHEISTRDSALODTDDSDPVLIGARIAPGDRSAVA 681

Db 553 ----- 552

QY 682 RIQEFRRKERMKEELDTLIRPLVKYKGRNSRTIMEANFNG--ANFWSGSD 739

Db 553 -----YKHINSRT-IKGVNFEFGPKCEYIVSSD 580

QY 740 CGHFIWDRHTAEHLMLADN-HVNCLOPHFDPILASSGIDYDIKIMSPLESRIFN 798

Db 581 CGNIFFWDKNTAEIATINMKGDHGVNCLPEHPMVPYATSGLEHDVYKIMTPGGERKLP 640

QY 799 RKLADVEYITRN-----ELMEEFTRNTI-----TVPASFMILMASLNHTPADKLE 843

Db 641 EDLKLQTLQRLFRFCNIVESIDLINNFQYTRGLGSGASHLRQRTHTSLGHQMRDN-S 699

QY 844 GDRSESGOEN 854

Db 700 SSRNSGSMASN 710

RESULT 15

Q9W091 PRELIMINARY; PRT; 748 AA.

AC Q9W091; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE CG8001 protein.

GN CG8001.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceinliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,

RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,

RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Hoch S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischman W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jaisl M., Kalush G., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mekulov G., Milshina N.Y., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1 SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AE003472; AAF47563.1;
 DR FlyBase: FB00035268; CG8001.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 6.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat: WD repeat.
 SQ SEQUENCE 748 AA; 83334 MW; AC92C5F0E3ABFC1A CRC64;

Query Match 8.2%; Score 373; DB 5; Length 748;
 Best Local Similarity 17.3%; Pred. No. 4, 9e-16;
 Matches 147; Conservative 73; Mismatches 163; Indels 468; Gaps 14;

QY 31 YLGRREFQRLKLEATLVNHDGCVNTICWDTGTYLISGSDTKLYISNPKVLTTR 90
 DB 306 YGSRQVQVQWTLSSLVNHGCVNLSNFRAGDLCISGSDLTIVVDMKKEQLHFR 365
 QY 91 SGHRANIPSAFLPCTNDKOIVSCSGDVIFYTNEQDAETNRQOQFCHGTTEIMTV 150
 DB 366 SGHNNINIFQTFIDSAGLDIVSSRRQVRSVYPSGVIKPKRLTHSESVKILIV 425
 QY 151 PNDPYTSLCGEDGTVMFDTIRIKTSCTREKCDKDLINC-----RRATSVACIP 201
 DB 426 PSHRHEILMSAGEDAAVYKFLRASNAAT-----TMMRCYNDENENGVRFLSLAHNP 478
 QY 202 PIPYLVAGCDSVRIYDRMLGTRATGNVAGKGTVMARFIPSHLNKSCRYTSLCY 261
 DB 479 YAPFCVSG--SDILRYDKRNL--AKALHOMAPRNL-----LEAOITQITCAVY 525
 QY 262 SEDQELIVSYSDYIYLFDPKDDTARELKPASAEERREELROPVYKRLRGMSDGP 321
 DB 526 NHHSEILASVSDAGITLFDNRNN-----RGD----- 554
 QY 322 RAPESERERDGEOSPUNVSLMQRMSDMLSRWEEASVEVAOSNRGRSRRPGTSSDIS 381
 DB 555 ----- 554
 QY 382 TLPTVPSSPDLVSEVETAMEVDTPAFLQPSSTSMASQAHSTSSPTSPHSTPLSSPD 441
 DB 555 ----- 554
 QY 442 SEQRQVSEAGHHTHQSNDNNNEKLSPPKGTGEFVLSLHYSTEGTTSTIKLNTDEMSS 501
 DB 555 ----- 554
 QY 502 IASSRKIGCHCKSGEGESFVPOSSVQPPGDSERKAPESSSDVTKYOGVSAENPVE 561
 DB 555 -----LHC----- 557
 QY 562 NHIIITQSDKFLAKPLDSNGERNDLMDRSCGVPEESASSEKAKEPETSDQTSATN 621
 DB 558 ----- 557
 QY 622 ENNTNPEQFQTEATGSAHEETSTRDSALQDTPDDDDPVLIPGANYRAGGDRSAYA 681
 DB 558 ----- 557
 QY 682 RIQEFFRRRKRKEMBELDTLNIIRPLVKVYKGRHSRTMIKEANFWG--ANFVMSGD 739
 DB 558 -----YKGINSRF--IKGVNFEGRSEIYVSGSD 585

QY 740 CGHIFWDRHTAEHMLLEADN-HVNCIQPHFPDEIILASSGDYDIKIMPSLESRIN 798
 DB 586 CGNIFPMQKNTAIIINMGDVGAVNCEPFPMPVLTSLGLEDVKTWTQGPCKLP 645
 QY 799 KRLADEVITRN-----ELMEETNTI-----TVPASFMRLMASLNHRADRL 843
 DB 646 EDLKTQTLRNRNCNIVESIDDINNFGYIFNGFLOGSGASHLRHTSTSLGHQMDN-S 704
 QY 844 GDRSGSGGEN 854
 DB 705 SRSNGSNASN 715

RESULT 16
 Q9NV87 PRELIMINARY; PRT; 615 AA.
 AC Q9NV87
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CDNA FJ10872 f1s, clone NT2RP4001725, weakly similar to guanine
 DE nucleotide-binding protein beta subunit.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Negahari K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AK001734; BAA91868.1;
 DR InterPro: IPR001005; MYB_DNA_binding.
 DR InterPro: IPR001440; TPR.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00515; TPR; 2.
 DR Pfam: PF00400; WD40; 4.
 DR ProDom: PD000018; WD40; 1.
 DR SMART: SM00320; WD40; 3.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 2.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat: WD repeat.
 SQ SEQUENCE 615 AA; 69049 MW; D5930944AB691PF CRC64;

Query Match 7.3%; Score 330; DB 4; Length 615;
 Best Local Similarity 24.3%; Pred. No. 2, 4e-13;
 Matches 144; Conservative 79; Mismatches 275; Indels 94; Gaps 21;

QY 13 DYKRSIGLEDPSSRLRSRYLGRREFIOLKLEATLVNHDGCVNTICWDTGTYLISGSD 72
 DB 9 DLIRQIKERGALSFEFRHYHTDPFIRRLGLEAELQHGSCVNCLENNKEKDDLASSGD 68
 QY 73 TKIYISNPKVLTTRISGRANIPSAFLPCTNDKOIVSCSGDVIFYTNEQDAETN 132
 DB 69 QHTVMPDLHHKLLSMHTGHTANIFSVKFLPHAGDRLLTGADSKVHYDLTVKETHI 128
 QY 133 ROCQFTCHYGTTVEIMVNPDPYTFILSCGEDGVATWMDTRIKTSCTYKDDK--DDILTN-- 189
 DB 129 ---MGDHTNRNVRKIATAPAMPNPTFWASAEDGLIRQYDLR-----ENSRHSEVLLDLT 178
 QY 190 ---CRR---AATSVACIPPIPYLLAVGCSDSVRIYDRML-----GTRATGNVAGRG-- 237
 DB 179 EYCGQLVEAKCIIVNPDDNCLAVGASGPFVRLDIDIMHNRKSMQGPSAGVHTFCDR 238
 QY 238 ---TGAVARFIPSHL-----NKKSCRY---TSLCYSEDGQETIVSYSDYIYFDD-- 281
 DB 239 QKPLPDGAOYVAGHLPVKLPDYNRLRVLVATYVTFSPNGETLILNMGGEGVYLDLT 298

QY 825 SFMLKMLASLHNRADRLGRCRSGSGOENENDEE 860
 DB 316 DPLEVMLLMNGY---RITGLSGAGASDDEDSSE 347

RESULT 19

ID Q96JK2 PRELIMINARY; PRT; 958 AA.

AC Q96JK2; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE KIAA1824 protein (Fragment).

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxId=9606;
 RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;
 RX MEDLINE-21245130; PubMed-11347906;

RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.

RL DNA Res. 8:85-95(2001).
 CC EMBL; AB058727; BAB47453.1;
 DR Interpro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PROSITE: PSS0082; WD_REPEATS_2; 3.
 DR PROSITE: PSS0294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.

FT NON_TER
 SQ SEQUENCE 958 AA; 105178 MW; CF6799191C9F7B6E CRC64;

Query Match

Best Local Similarity 5.18; Score 232.5; DB 4; Length 958;

Matches 170; Conservative 112; Mismatches 343; Indels 203; Gaps 32;

QY 16 KRSLGLEPRLSRVYGR-----EFIRKLEATLVN-----HDGCVNITCW 59
 DB 18 KRRAGLGSGMRVVGFLSQRGLHDDPLLTDPQRRIRGCRNLKTKLLGHFGCVNAIEF 77
 QY 60 -NDTEYILSGSDTKLVISNP---YSRKYLTIRSGHRANIF-----SAKFLPCTN 107
 DB 78 SNNGGQWLVSQGDRLRVLLMHMEQALHSRYKPIQLKGEHSHNIFCLAFNSGNTKVFSGGN 137
 QY 108 DKQIVSGSGGVLFYFNVEDDAENRRCQCTCHYGTTEIMVYNDYITLSCGEDGTVR 167
 DB 138 DEAYIL-----HDVESSEFLDYFAHEDAVYGLSVSPVNDNITASSDDGRVL 184
 QY 168 WFDTRITKCTKEDKDDILINCRATSVAIAPPIYVAVGSDSVIRYDR----- 222
 DB 185 IWDIR-ESPGEPC-----LANPFAHSHVMPNVEPRLLATANSKRGVGLMDIRKQSS 239
 QY 223 -----MGTRATGVYAGRTTGM-VANFIDSHLNN-----KSCRV 257
 DB 240 LIRYVGNLSLQASMSVAFNSGTOLLALRRRLPVIYDISRLPVFOFQOGYFNISCTMK 299
 QY 258 SLCYSEDGRIIVSYSDY-IYLFDP-KKDTAREL-KTPSAERRELRQPPVKRLRLRG 314
 DB 300 SCCFAGDRDYILSGSDDFELWYMRIPADPEAGIGVYVGAFWVLKGHSITVQVAFNP 359
 QY 315 D-----WSDGPRAPRESEREREQSPNVLQMRMSDLRMFEASEV 359
 DB 360 HTYMWISSGVEKLIKIMS---PYKPGCTDDLDGRLEDRCILYTHETYS----- 407
 QY 360 AGSNRGRGRNRPGRGTS-QSDITLPTVSSPDLVSEIAMEVDTPAEOFLPSTSTMS 418
 DB 408 LVYLSGSGLSHDYANOSVQEDPRMAFFDSLVRREIEGMSDSDSD-----LSESTILQLH 463

QY 419 AQAHSTSPTESPHSTPLLSPPDSEQRQSVSEASHHTHOSDNNNEKLPKRGTEPYLS 478
 DB 464 AGVSRSGSYTSESSASLPREPP-----PYDESADNAH-----LGP----- 501
 QY 479 LHYSTEGTTITIKINTDEMSS-----IASSRIGSHCKSGOE 520
 DB 502 LRVTTNTVASTPPPTCEDAASRQRLSALRRYQDKRLALSNRSEENVECEVETD 561
 QY 521 SFVPOSSVQPPGESETAPESSEDYTKYQGVSA---ENPVENHITQSDKTA--K 575
 DB 562 LF-FRPRSPSEDESSSSSSSEDEELERRASTWORAMRRRQRTTREDKFSADIK 620
 QY 576 PLDSNGERN---DLNDRSCGVPEESASSEKAKEPETSDOTSEATNENNTPEQF 631
 DB 621 PNTYIGEDNDYDPIKYND-----DLSSSPISPRSTL-----EI 658
 QY 632 QTEATGPAHEETSTRD-----SALDDTDSDDPYLIPGA---RYRAG----- 672
 DB 659 QPSRASPTSDIESEVERKLYKAYKMLRYSYISYNNKDETSLVGGEADGCRAGSHKDP 718
 QY 673 -PGDRASVARIQEFERRKKEKMEELDTLIRNRLYKMYKGRHS 719
 DB 719 APSSKKECLINAAQRNODLPPEGCKSDTKETPTPSNGPHEHS 766

RESULT 20

ID Q9V5L7 PRELIMINARY; PRT; 747 AA.

AC Q9V5L7; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE CG12892 protein.

GN CAP1-105 OR CG12892.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidae; Drosophilidae; Drosophila.

ON NCBI_TaxId=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Chapple M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Esler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris J.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
 RA Jaitani M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlynshina N.V., Mobarry C., Morris J., Moshirei A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazon M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AE003829; AAF58788.1; -
 DR Flybase: FBgn0033526; Cafl-105.
 DR InterPro: IPR001632; GpoteIn_B.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PRINTS: PR00319; GPROTEINB.
 DR ProDom: PD000018; WD40; 5.
 DR SMART: SM00320; WD40; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; 1.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 DR Repeat: WD repeat.
 KW Repeat: WD repeat.
 SQ SEQUENCE 747 AA; 83394 MW; 0CF0BA42FCA2D89D CRC64;

Query Match

Best Local Similarity 4.7%; Score 213.5; DB 5; Length 747;
 Matches 167; Conservative 103; Mismatches 282; Indels 257; Gaps 40;

OY 8 PHLL-----DYKRSKSLGDEPS--RLNSR-----YLGRREFIQRLKLE--A 45
 DB 6 PEISWNRDPLVSDIQNGILGRSPTRICRLASGSDAHLYIVNRSDAEVDIELA 65
 OY 46 TLNVHGCYNTICMNDTGEYIISGSDTKLYT-----SNPSRKVLT 88
 DB 66 DISRHQRAVNAWVSPNGELASGDSESVYIWKQADHEVNIIVADGSCSEODKEWILT 125
 OY 89 IR--SGHRANIFSAKFLPCGNDRKOIYSGSDGVIFTYNVEQ----- 127
 DB 126 LKYLBRHREDIYDLNMAP--NSQFLVSGSVYDNTAMLMVDHSGSLAIIIDHKYVOGVAM 133
 OY 128 -----DAETNR-----OCQFTCHYGVTEYIMVTPNDPYTF--LS 159
 DB 184 DPCNQIATMTSDRQMRIFDANKRVILHRYSK-----VLPVKDHMHGKSMR 232
 OY 160 CGEDGVNRFDRIKTSCKRECK-----DIL--IN-----CRAATSAVACP 201
 DB 233 LYODGTLQFFRRL--CFTPDCKLLTPSGIYDGVVAKPINTSYGFSHHDLSKPAFVL 289
 OY 202 PIP--YLLAVGSGDSSVRL--YDR-----RMLGFRANGYAGRTTGAVARF-I 245
 DB 290 PFPKEAVAVRGSFVYLRLEPNAEKNPPLISLPYRMITAVATKNAVFFYDTQOPVFAI 349
 OY 246 PSHLNKSCRVTSIACYSEDQELIVSYSDYIYL--FDPKD--DTARELKT-----SAR 296
 DB 350 VSNIHYS--RLDPLAMSSDGVLYIVSSTGYCGLIFFEPELGDCEYEDMETVLSVYLKSS 407
 OY 297 ERREELRQPPVKRLRLGWSMTGPRAPREBERDERDEOSPVSILMQRMSDMLSRMFEA 356
 DB 408 ENATYVLRK--KROKLRKSLDE-PR-KPELOEKS-----PNT-----IRRA 444
 OY 357 SEVAOSNRGRSRPRGSGSDISFLPYVPSDELEVESETAMENVDTAPAQFQPTSSRT 416
 DB 445 SEAGITEVEGEPELDAENDSSTHSVSNKT--NSPKTKASEE--EKPTPLAIRSRPKNST 500
 OY 417 MSAQASHSTSPESPHSTPLSSPDEOROSVASGHTHHSODNNNEKLSPPKGTGEPEV 476
 DB 501 -----APMPILIRAPRKP-----EDKNGSESKRPD----- 527
 OY 477 LSLHSTEGTITSTIKLNTTDEMSSIASSRGISCKEAGEESFVPOSSVOP----- 531
 DB 528 -----EMEYTHRTQVOYKT-----VASPVKRVSTEAVPAPETSPALAVIP 570
 OY 532 -----EEDSETKAPESSESDVTKYQEGVSAENPVENININITOSKFTNAKPLDSSNGR 584

DB 571 VFEKEITSSDDKEESPCKSRPATPIYVRQOPRT-----GSSQFNITPKSQPAKQA 623
 OY 585 NDILNRSCGV---PREES--ASSEKAREPETSQDSTE--SATNENTNPEQOFQTEATG 637
 DB 624 TPVAVRRTPVLIEMPVNPVPEEAMDPALDESTPLPATVKKDSRKLPLPVKKEIK 683
 OY 638 PSAHEETS-----TRSDALQDITDSDDDP 661
 DB 684 PAVTEASEATCERTEDIRLYVEDTQEEETP 712

RESULT 21

ID O95TD6 PRELIMINARY; PRT; 298 AA.
 AC O95TD6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE SD07783p.
 GN CG5124.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Phrygotea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AY059458; AAL13364.1; -
 DR Flybase: FBgn0034305; CG5124.
 DR InterPro: IPR000923; BlueCu_1.
 DR InterPro: IPR001440; TPR.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 2.
 DR PROSITE: PS00196; COPPER_BLUE; UNKNOWN_1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat: WD repeat.
 SQ SEQUENCE 298 AA; 33989 MW; C5A54507F669A9A43 CRC64;

Query Match 4.7%; Score 212.5; DB 5; Length 298;
 Best Local Similarity 40.2%; Pred. No. 4e-06;
 Matches 51; Conservative 24; Mismatches 39; Indels 13; Gaps 4;

OY 693 RKEMEELDTINIRPLVKKYKGRNSRMVIRKANFNGA--NRYMGSGDCGHFIPIIDRHT 750
 DB 157 RLKNAEYDLRSTARDYQR--YVGHCVTTDICKRANLYIGSGEFLAAGSDGNNYIMWGDY 215
 OY 751 AEHLMLLEADNHYVNCLOPAPFPDPIILASGIDYDIKIMPLESRJPNRKLADQVITRNE 810
 DB 216 GKIRAYRADSAIVNCVQPHPSICMLATSGIDINIKIMSCAAS-----AEE---RPN 265
 OY 811 LMLEETR 817
 DB 266 LVADVTR 272
 RESULT 22
 ID O9BIT5 PRELIMINARY; PRT; 747 AA.
 AC O9BIT5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
 DE Chromatin assembly factor-1 p105 subunit.

GN CAF1-105 OR CG12892.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RX MEDLINE=21424640; PubMed=11533245;
 RA Tyler J.K., Collins K.A., Prasad-Sinha J., Amiot E., Bulger M.,
 Harter P.J., Kobayashi R., Kadonaga J.T.;
 RT "Interaction between the Drosophila Caf-1 and ASF1 Chromatin Assembly
 Factors";
 RL Mol. Cell. Biol. 21:6574-6584(2001).
 CC -1 SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AF367178; AAK31264.1; -;
 DR Flybase: FBgn0033526; Caf1-105.
 DR InterPro: IPR001632; GpoteIn_B.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PRINTS: PR00319; GPROTEINB.
 DR ProDom: PD000018; WD40; 1.
 DR SMART: SM00320; WD40; 5.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 747 AA; 83406 MW; C7EBE83924388767 CRC64;
 Query Match 4.7%; Score 211.5; DB 5; Length 747;
 Best Local Similarity 20.6%; Pred. No. 1.6e-05;
 Matches 167; Conservative 103; Mismatches 282; Indels 257; Gaps 40;
 QY 8 PHILW-----DYKRSLSGLEDS--RLNSR-----YLGRREFIQRLKLE--A 45
 DB 6 PEISWMNRDPVLVSVDIQNGIGIRSPTRICRIASGSDAHVLIWYVRSDABEVDLEIA 65
 QY 46 TIANVHDCVNTICMNDTGEYIISGSDTKLYI-----SNPYSRKVLTT 88
 DB 66 DLSRHQRAVNAVWSPNGELIASGDESVYFIMKORADHEVNIYDADGSEQDKVWLT 125
 QY 89 IR--SGHRANISAKPLPCTNKOIYSCSGDGVIFTYNEQ----- 127
 DB 126 LKYLKSHREDIYDLWAP--NSQFLVSGSVNDNTAMLVHSGKSLAIIIDHKCYGVGVAM 183
 QY 128 -----DAETNR-----OCQFTCHYGYEIMTVNDPYTF--LS 159
 DB 184 DPCNOYIATMSTDRQMRIFDANTKRVLHRSKC-----VLPVKEDHMHGSKMR 232
 QY 160 CGEDGVREDFRIKTSCKRECK-----DDIL--IN-----CRAATVAICP 201
 DB 233 LYQDGTLOTFEERL--CETPGKLLLPSTGYDGVVKKPINTSYGFSKYDLSKPAFVL 289
 QY 202 PIP--YLLVAGSSDSSVR--YDR-----BMLGTRATGNYAGKGTGMVARF-I 245
 DB 290 PPPNEAVAVKRCSPVLYRLRPYNAEKNPILSLPYMIITAVATKNAVFFEDIDQPPPAI 349
 QY 246 PSHLNKSCRVTSLSYSEDGQELIVSYSSDYTL--FDPKD--DTARELKT-----SAE 296
 DB 350 VSNIHYS--RLNDLWMSDQGVLIYSTDQYCSLITFEPELTDCEYEDMETVLSVLKSS 407
 QY 297 ERREELROPVRLRLRGCMSTGPPARPESEEREDEGSPNVSLMQRMSDMLSRFEER 356
 DB 408 ENATYVKK--KROKLKRVSLDE-PR-KPLQEKSK-----PNT-----IRRA 444
 QY 357 SEVAOSNRGRGSRPRGSGQSDISLPLVPVSSPDLEVEETAMEVDTPAEOFIPSTSS 416
 DB 445 SEAGTIEVEPELDANDSTSHSVSNKT--NSPKTKASEE--EKPTPLAIRSPKKNST 500
 QY 417 MSAQAHTSSPTSPSTSLSSPDEQRQSVASGHTTHOSDNNEKLSPPKGTGEVY 476
 DB 501 -----APMPAIRARAPRP-----EDKNGSESSKPPD----- 527

QY 477 LSLHSTEGTTSTIKLNFTEWSSIASSRGICSGCKEGBQESRPVQSSVQP----- 531
 DB 528 -----EMEYVHTKTTOVOKT-----VASVKKRVSTEARPAETSSQPALAVIP 570
 QY 532 -----EGDSEIKAPEESEDVTKYQGVSAENPENNINIQSKPFAKPLDSNGER 584
 DB 571 VREKELISSDKFESEPKSKRPATPIQVRROPRT-----GSSQFNLTPKSPAKQA 633
 QY 585 NDNLDRSGCV-----PEES--ASSEKAKEPETSQDTSTE-SATENNNTPEPOFQTEATG 637
 DB 624 TPFAVARTPRVLEIMVNSVPVMEAMAMPDESITPLPATKKSRRKPLPTVTEFK 683
 QY 638 PSAHEETS-----TRSDALODTDDSDDDP 661
 DB 684 PAVIEASEATCETEDIRLYVEDTQETP 712
 RESULT 23
 ID Q9FWE0 PRELIMINARY; PRT; 131 AA.
 AC Q9FWE0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Putative WD-repeat containing protein.
 GN OSJNB001511.22.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ebrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Bueli C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsu J., Zismann V., Pei G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNB001511 genomic sequence";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AC051633; MAG13587.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 2.
 DR ProDom: PD000018; WD40; 1.
 DR SMART: SM00320; WD40; 2.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 131 AA; 14750 MW; EA37C0E1C19A32F9 CRC64;
 Query Match 4.6%; Score 207; DB 10; Length 131;
 Best Local Similarity 35.5%; Pred. No. 3e-06;
 Matches 44; Conservative 25; Mismatches 51; Indels 4; Gaps 1;
 QY 1 MSKGSYF-----HLMDVKKRSLSGLEDSPLRSKRYIGRRFIORLKLEATLVNHDGCVNT 56
 DB 1 MRPPWHPKARKAGADLCIREVGLLPFRFARRAASEDLVRLQVHRLNHTGCVNT 60
 QY 57 ICNNDGELYISGDDTKLVISNPKRYLTTIRSGHRANIFSAKFLPCNDKOIYSCSG 116
 DB 61 VGFNADGDTLISGDDQVLMWDMGTGAIKLPQSGSHSNVFPARRPPTNDQITVCA 120
 QY 117 DGVI 120
 DB 121 DGEV 124
 RESULT 24
 ID O60336 PRELIMINARY; PRT; 1217 AA.
 AC O60336;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE KIA00596 protein (fragment).
 GN KIA00596.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AB011168; BAA25522.1; -;
 DR Interpro: IPR002114; HPr_Serp_site.
 DR Interpro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 8.
 DR PRINTS: PR00320; GPROTEINBPT.
 DR SMART: SM00320; WD40; 7.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 DR PROSITE: PS50082; WD_REPEATS_2; 1.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 FT NON_TER
 SQ SEQUENCE 1217 AA; 131098 MW; 0BC4EAC6722BEF5 CRC64;
 Query Match 4.5%; Score 204.5; DB 4; Length 1217;
 Best Local Similarity 20.6%; Pred. No. 8,76-05;
 Matches 187; Conservative 121; Mismatches 351; Indels 247; Gaps 43;
 QY 39 ORLKLKTLNVDGCVNTICWN--DTGEYLISGSDPTKLV---ISNPKRVLTIRSG 92
 DB 218 EMLKVEA---HSEILCLESKRPDTGKLILASNRRLIHVLADAGREYS---LQOTIDE 270
 QY 93 HRANIFAKELPCTNDKOIVSCSGDVIYFTNVBDAETNQCQFT-CHY---GTYEI 147
 DB 271 HSSITIAVKFAASDGVMISCGADKSIYFTAKSGD--GVQFTRHVVRKTLTYDM 327
 QY 148 MYVNDYFTLSCGEDTGMFDRIRKTSCKEDCK---DILLINCRRAATVAILCP 202
 DB 328 DVPSKMTYIIGC-QDNINILFN--ISSGKQKLFKSGGEDGLIVQ-----TDP 376
 QY 203 IPYLLAVGSDSSVRIYDRMLGTRATGNTAGRTGMVAFIPSHLNKSCRYSLCYS 262
 DB 377 SGYIATNSCDKMLSTFD-----FSSGECVATMTGH---SEIYTKKFS 417
 QY 263 EDGQELIVSSP-YIYLFPPKDDTANELTPSAEERREE---LRQPPVKRLRLG--- 314
 DB 418 NDKKH-LIYSVGSCLFVWRLISEMTISMRORLAELRQRGKQGPSPQSRASGNRH 476
 QY 315 ---DMSDTPGRAPRESEREDG---EOSPVSIMQR-----MSDMLSMWE 354
 DB 477 QASMLSPGALSSDSKDEDEGTEDLPALPYLAKSTKALASVPSPALPRSLSHW- 534
 QY 355 EASEVAOS-----NNGRGRSRPRG-----GTQSD 379
 DB 535 EMRAQESVGFLLPAPAPNGPRRRGRWVQGYELSVRMLDLRLQTLTAPSLQDPQDS 594
 QY 380 ISLLPYVPSPDLEVTAEVITPAQFLQPTSSSTMSQANSTSPSPSPHSTPLSS 439
 DB 595 LAITPSGPRKHGQEALETSL-----TSONEKPPRQASQPCSYPHIRLLSQ 641
 QY 440 PDESOROSVEAS---GHHTHOSDN-----NNEKXSPK 469
 DB 642 EEEGVFADLEPAILEDGIYVPEPSDNPMTSEFOVQAPARKGLGRYPPSSRSSEKSP- 700
 QY 470 PGTGEVPLSLHYSTEGTSTTIKLNFTDEMWSIASSSRIGSHCK--SEGQESFVPOSS 527

DB 701 ---DSACSVDS--SCLSPEHPTDESESTEPLSYDGISSDLPEAPAGDEEEEEE 754
 QY 528 VQP---PEGDESKAPEE--SESDVTKYQGVGAENPV---ENHINITOSDKF-----TKP 576
 DB 755 MCPYLOQESRPTPOPOEOLKHFETLASGAPGAPVQVPESESRSISRRLLQVQTRP 814
 QY 577 LDSNGERNLNL-DRSCGVPEESASSEK---AKPEPSTDOTSTESANTENNTNPEQ-- 630
 DB 815 LNEPSSSSILALMSRPQAVQASGEQPRGNCANPPGAPPEVPESSG-----NNSPQA 868
 QY 631 -----FQETAGPFAHETSTRDNLQDTDSDDDPVLIIPARRA-----GPDR 676
 DB 869 ASVLLPRRLNDPSSWAPKRVATASPFGLQAQSVHS--LVQERHESLQAPSPGAL 925
 QY 677 RSAVARIOEFFRRKRRKEMEELDTLNIIRPLVKWYKGH-----RNSRTWKANPWG 730
 DB 926 LS-----KEIEAQDGLSIPADGPPSRHSIQNTTSMATKSISYV 970
 QY 731 ANFWAGSDCGHIFTWDRHTAEHMLLEADNH-VVNCLOPHFPFDILASSGIDYDIKIMS 789
 DB 971 ENLGLVAPPOAHAPI--HVSPLSKLALPSRAHLVLDIPKPLDRLTLA-----FS 1019
 QY 790 PLEESR 795
 DB 1020 PVTGR 1025
 RESULT 25
 ID Q17490 PRELIMINARY; PRT; 5170 AA.
 AC Q17490;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE B0350.1 protein.
 GN B0350.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopira A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gattung S.;
 RT "The sequence of C. elegans cosmid B0350.";
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U50071; AAA93447.1; -;
 SQ SEQUENCE 5170 AA; 575848 MW; 27C2DFBFAE03A7B CRC64;
 Query Match 4.5%; Score 202; DB 5; Length 5170;
 Best Local Similarity 20.7%; Pred. No. 0.00089;
 Matches 125; Conservative 84; Mismatches 248; Indels 146; Gaps 24;

DB 558 GGGST-----TSSDPVEASGTCNGN-----SSTOSSSTTTTSSDDEQTTSS-SSDPV 606
 QY 593 CGVEBSASSEKAREPETSQDSTESATN-----ENNTNPEPQTEATGSAHEETSTRD 648
 DB 607 SEVAGSSSSIGDGNSSGSSSTTTTSSDGGOSTSSDP--VEVAGSGTNGNSSTOS 664
 QY 649 SALQDTDDSD-----DPLTPGARYPAGPDRRS 678
 DB 665 SSSSTTTTSSDDEQTTSSDPV-VEVAGSSSSNGDGN 701

RESULT 27

ID 096006 PRELIMINARY; PRT; 1183 AA.
 AC 096006;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE LD47622P.
 GN CG7337.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,
 RA Gonzalez M., Garin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacle J., Paragas V., Park S., Phuanenavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AY052025; AK93449.1; -
 DR Flybase: FBgn0031374; CG7337.
 DR Interpro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 7.
 DR PROSITE: PS50082; WD_REPEATS_2; 2.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 1183 AA; 128147 MW; 43B043CE7A870BCD CRC64;

Query Match 4.4%; Score 200; DB 5; Length 1183;
 Best Local Similarity 20.3%; Pred. No. 0.00016;
 Matches 153; Conservative 107; Mismatches 333; Indels 162; Gaps 28;

QY 41 LKLEATLNVHDCVNTICNDTGEY-----LSGSDDTKLVISNPKYVLTIRSGHRA 95
 DB 176 LRLTTIEAHES--EVLCEYSNEKIERKLLASASRDRLHFDVAVQNYLLQLTLDHSS 233
 QY 96 NISAKFLPCTNDKOIVSCSGD-----GVIFYNVBDQAEFNQCGFTCHGYTTE 146
 DB 234 SITSIKIFVAGLNFQMIISCADSIKFRSGNIFMGTSCT-----TLVD 282
 QY 147 IMTPVNDPYFLSCGEGDGVYRWEDTR-IKTSCTEKDCKDILINCRRAATVAICPPIPY 205
 DB 283 MEVDSNAKHLLTAC-QDRNRYVGTQNAKQTKFKGSHD-----EGSLIKLSLDPSGI 335
 QY 206 YLAVGSDSSVRYD-----RRLMGRATGNAGGTTGWARFIPSHLNKSCVYTS 258
 DB 336 YVATSTCKTLAVDYYSNECMARMG-----HSELVTG 369
 QY 259 ICSYSDGQELIVSYSDYILFPKDDTARELTPSAEERRELR-QPVRKRLRLGDM 316
 DB 370 LKTNOCRHILISAGGICIFIMQVPHDM---ITTMQARMQQLRSHAPLPRPLADISP 426
 QY 317 SDTGPRAPRERERDEGSPNVLQRMDSM--LSRMFEASEVAQSNRGKSRPRGG 374
 DB 427 PDGIVLESPTSEIOP-QLOPKFGVAERFSDVQQLPQWAMARKAAADSDGSLSTPTPSGC 485

QY 375 TS-----OSDITLPTVPSSPDLEVESTAMEYDTPAEQFLQPTSTSTMSAQHSTSPTES 430
 DB 486 SATVPMGMAAASMGONLSSPSQMTGLAPRARGMWQ-----RSTQLETRADDRSNSES 539
 QY 431 P-----HSTLLSSPDEQROSVASGHHTHHOSDNN-NKLSKPKPTGTPVLSH 480
 DB 540 PLGTGVSSVGGSHGVNVOVTSDNSASSKDTYNTQYLESDSSIDSGMTRRELAFIGSSN 599
 QY 481 YSTEGTTTSTIKLNFTEWSSIA-SSSRGIGSHCKSEGESEFVQSSVOP-----530
 DB 600 NGTVVTVSV-----SSINVASNGAMSTSGGAQQLQLPDKRKGLKGLFDTHH 650
 QY 531 -PEGDSETKAP-EESSEDVTKYQEGVAENVEHNINTQSDKFTAPPLDSNGERDLN 588
 DB 651 DHDGDVEDISDGETRSSDHGMFNPLAPSPPTDKVTAMNDEL-RKSVRQKREKSLQ 709
 QY 589 LDRSC--GVPEBSASSEKAREPETSQDSTESATNENTNPEPQTEATGSAHEETSTRD 646
 DB 710 LTPSALSGNGSSHTASTGTSTDEKSTPSA-----ENKERSLASTLGSSSNLPQSS 764
 QY 647 RDSALQDTDDSDDPVLPGARYPAGP-----DRAS-----AVARIQE 685
 DB 765 TNSFLH-----AALPEGGLTTPMERGSSRSISAKHTENGKSVAPPT 810
 QY 686 FRRRRKRRKMEELDTNIRRLPKVYKGRHSR 720
 DB 811 ITRKYSTKKEELLQVINKYQOLENV--GHRPLR 843

RESULT 28

ID 09V054 PRELIMINARY; PRT; 2351 AA.
 AC 09V054;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE CG7337 protein.
 GN CG7337.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garcia N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris R.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Iregyan C.,
 RA Jialili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spelling A.C., Stapleton M., Strong R., Sun E.,
 RA Svetskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000)
 CC -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL: AE003584; AAF51327.1; -
 DR Flybase: FBgn0031374; CG3337.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 10.
 DR SMART: SM00320; WD40; 9.
 DR PROSITE: PS50082; WD_REPEATS_2; 2.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat.
 SQ SEQUENCE 2351 AA; 258254 MW; AE1402C714709AA8 CRC64;

Query Match 4.4%; Score 200; DB 5; Length 2351;
 Best Local Similarity 20.3%; Pred. No. 0.00042;
 Matches 153; Conservative 107; Mismatches 333; Indels 162; Gaps 28;

QY 41 LKLEATLVNHDGCVNTICWNTDGEYI-----LSGSDTKLVISNPKYVLTIRSGHRA 95
 DB 501 LRLITLIEAHES--EVLCELYSNEKIERKILASARDRLIHVDVANOYLLQTLDDHSS 558
 QY 96 NIFSAKFLPCTNDKQIVSCSD-----GVITYTNEQDAETNRCCOFTCHYTYE 146
 DB 559 SITSIIFVAGINLFQMIISGADISIMFRSFGNIFMRKGTSTGKT-----TLYD 607
 QY 147 IMVVPDPTFLSGEDGVYRMEDTR-IKTSCKECKDKDILINCRRAATVAICPPIPY 205
 DB 608 MEYDSNAKILTRAC-ODNRNRYGTQNAQOTKFKGSHSD-----EGSLIKLIDPSGI 660
 QY 206 YLAVGSCDSVRIYD-----RRLGTRATGYNAGRITGVARFIPSHLNKSGRTYS 258
 DB 661 YVATSTGDKTALAYDYYSNMCAMRMG-----HSELVYG 694
 QY 259 LCYSEGGELIVSYSSDIYILFPPKDDTARELTPSAERRELR--QPPVKRLRLRGDW 316
 DB 695 LKFTNDCRHLISAGDGCIFIMQVPHDM--IYTMQARMSQGLRSGHAPLPPLAPISP 751
 QY 317 SDTPRARESERERDEGSPNYSMLQRMSDM--LSRMFEAESEVAQSNNGRRSRPRG 374
 DB 752 PGIVLESTPTELEOP-QIQPKFVAREKSDVGQLPOMARKKAADSDGALSIPPSGG 810
 QY 375 TS---QSDISTLPVPSPPDELEVSETAMEVDTPAEQLOPSTSTMSAQAHSTSPTES 430
 DB 811 SATVPMHAASSKGNLSSPSQOMTGLAPRARGMAQ-----RSQULEADADLRNSSES 864
 QY 431 P-----HSTPLISPDSEGRQSEVSGAHTTHQSDNN-NEKLSPKPGICEPYLSLH 480
 DB 865 PLGTAVSVGSHGQVNVQTDYNSASSKRDITYNQTYLSSEDSIDSGMETRGELKFGSSN 924
 QY 481 YSTEGTTSTIKLNTDEMSTIA-SSSRGIGSHCKSEGOESPVPSSVOP-----530
 DB 925 NGVAVVAVSV-----SSTAVASNGAMSTGSCAAQORQLPCKRLKPGLRDPTHTH 975
 QY 531 -PEGDESLAP--EESSEDTYKQEGVSAENPVENHINITQSDKFTAKPLDLSNGERDNLN 588
 DB 976 DHGDVEDLSDGERTSSDGMRYNNLAPSTPDTFKYLANEDL-RKSVRQKQFEKSGIQ 1034
 QY 589 LDRSC--GYPEASASEKAKEPETSQDQSTESATNENNTNPEQFQTEATGPAHEETST 646
 DB 1035 LPPSALSAGSSSHIATGTGTDTEDEGSTPSA-----ENAEKSLASTLGGSSENLPQS 1089
 QY 647 ROSALQDTDDSDDDVPLIGARHAPG-----DRRS-----AVARIQE 685

DB 1090 TNSFLH-----AALPEGGLTTPMERGGSSRRSISAKHNTENGKVAAPPT 1135
 QY 686 FFRRRERKEMELDTLIRRPVKKWYKGRNSR 720
 DB 1136 ITKSYSTKKEELLOYINKKQOLENV--GHRRLR 1168

RESULT 29
 Q17343
 ID 017343 PRELIMINARY; PRT; 6994 AA.
 AC 017343;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-JUN-1998 (TREMblrel. 05, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE UNC-44 ankyrin.
 GN UNC-44.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RA Otsuka A.J.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U39847; AAB41827.1; -
 DR EMBL: U21733; AAB384.1; -
 DR HSSP: P42773; 11HB.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000488; Death.
 DR InterPro: IPR001360; GH.1.
 DR InterPro: IPR002383; GLA_blood.
 DR InterPro: IPR000906; ZUS.
 DR Pfam: PF00023; ank; 24.
 DR Pfam: PF00531; death; 1.
 DR Pfam: PF00791; ZUS; 1.
 DR PRINTS: PRO1415; ANKYRIN.
 DR PRINTS: PRO0001; GLABLOOD.
 DR SMART: SM00248; ANK; 21.
 DR SMART: SM00005; DEATH; 1.
 DR SMART: SM00218; ZUS; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 22.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 DR PROSITE: PS00572; GLYCOSTYL_HYDROL_FL_1; UNKNOWN_2.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Query Match 4.4%; Score 200; DB 5; Length 6994;
 Best Local Similarity 20.6%; Pred. No. 0.0018;
 Matches 124; Conservative 85; Mismatches 248; Indels 146; Gaps 24;

QY 191 RRAATVAICPPPIPYLAVGCSDSVRIYDRMLGTRATGNYAGRTGVAWAFIPSHLN 250
 DB 2344 RPISESPAVSERPSHATTTTTRVREYQDEKADQSSAKRES-----VRLVSEEH 2396
 QY 251 NKSQVTSICYSDEGQELIVSYSDIYILFDPKDDTA--RELKTPSAER-----REE 301
 DB 2397 HPACDSERLSEPAQSEPEVETHAESQFSNLYETTTTAAVREFYDDEQEOASRAATKER 2456
 QY 302 LRQPPKRLRLRDMSDGTGRAP--ESEREREGSPNYSMLQRMDSMLSRKFE--AS 357
 DB 2457 IEQSPVASER--SIVSTEHRQSPQESQSLSEPTSEKRVHTVETTTTTRVRECPEPIAS 2513

```

QY 358 EVAQNR--GRKSRPRGTSQSDISTLPVPSP--DLEVSEAMEYDTPAQFLDPTS 414
DB 2514 ELEHASREFEFGSGNDFRRSVPYDAOQYDVPHSPAPSSHAESOPVESPPIHIVETTT 2573
QY 415 STMSAASHSTSSP--TESPHSTPLSSPDSEORQ-----SVASGHHTHHOSNNN 463
DB 2574 TTVTRERODEEYRPPSP--AETFPPLSSQOSEPHIVAKETTTTTVTVELIDEPKGN 2631
QY 464 EKLSPRGT-----GEPVLS--LHYSTGTTSTIKLFTDEMSSIASSRGIGS 511
DB 2632 VTPSPAPSSHAESQVPEPVSSHQRYPHVETTTTTNTSNLYDEDDNVSEEDPATQ 2691
QY 512 HCSSEGESEFVPOS--STQPEGDS----- 535
DB 2692 HFQ--QSETSVHRSHDVSYESDDEGLGSKYLGFAKKAGVAGVYAAPYALAAGAKA 2748
QY 536 -----EKAPAESEEDVTAKYOE--GVSANENVENHINTQSDKFTAK 575
DB 2749 AYAPAEKDEDDETSHPESPVPYQSE--QYODDSAQSSHDFEHM----- 2794
QY 576 PLDSNGERNDLMDRSCGVPESSAKAKEPE--TSDQSTESATENNNTNPE----- 628
DB 2795 PESPHEKEETEEEDHS--HPESPVLSEKRDQVSETTTTTVTREYNDPEDEQEQ 2852
QY 629 -----PQOTATGPAHETSTRDSALQDTSDDPVLIPARAGGDRSAVAR 682
DB 2853 GPHSPAPSSHTAEHPHIVETTTTTVTREFQEEPE-----LEKQEDNSKSPSSH 2905
QY 683 IOE 685
DB 2906 SOE 2908

RESULT 30
ID Q23330 PRELIMINARY; PRT; 493 AA.
AC Q23330;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 53.6 kDa protein.
GN ZC449.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Latreille P., Gattung S.;
RT "The sequence of C. elegans cosmid ZC449.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41510; AAK39385.1;
KW Hypothetical protein.
SQ SEQUENCE 493 AA; 53643 MW; 6EDC2FEA3CCCA390 CRC64;

Query Match 4.4%; Score 199; DB 5; Length 493;
Best Local Similarity 19.3%; Pred. No. 5.9e-05;
Matches 102; Conservative 80; Mismatches 219; Indels 128; Gaps 16;

```

```

QY 249 LNKSCVATSLCYSEDOELIVSSDYITLFDPKDTAREK--TPSAREERELKOP 306
DB 59 LKNNKCLLTR-----ARRHLSTPATDDOSEETTT 91
QY 307 VKRLRLGDMSDTCGRPARPESERERDEQSPNVSLQMSDMLSRWEEA--SEVAQNRG 365
DB 92 VKSHKSKSTKTDPAKTEEPASTKKSEKKTEKKAVKTTTAPAPAESAFETPPSEEN 151
QY 366 RGRSPRGTSQSDISTLPVPSSPDLEVESEAMEYDTPAQFLDPTSSTMSAQHSTS 425
DB 152 KTEESGNVSAEEVTTTSTEVSEEQVYKSTEASTEEEK-----STASSEATTTTS 205
QY 426 SPSPSPSTPLSSPDSEQRQSVASGHHHTHODNNNEKLSPPKGTGEPLSLHYSTEG 485
DB 206 SEISE-----TTESEEE-----EETTPA-----KTHKSKT 233
QY 486 TTTSTIKLFTDEMSSIASSRGIGSHCKSEGESEFVPOSQVOPPEGSETKAPESE 545
DB 234 TTTT-----EASTKKDSKK--SHKTKKEKVTTPPESTESSSPSTSPSPQSE 283
QY 546 DVTYQEGV--SAENPVENHINTQSDKFTAKPLDSNGERNDLMDRSCGVPESSAS 603
DB 284 STPSEESTEQSTERP-----KKEDKKDKKKKKKKED-----SEEDDDK 327
QY 604 KAKREPISDOTSTESATENNNTNPEPQOTATGPAHETSTRDSALQDTSDDPVL 663
DB 328 KSKSSSSSDSKTDEKSTESDSSSEASEKTEVEGK-----KTLFSGPDEDDDD-- 379
QY 664 IPGARVRAGPDRRSAVARIOEFFRRKEREK---EMELDTLNIIRPLVKVYKGRHSR 720
DB 380 -----EGAGADAFSEKQATAPALSTRTTMAYKSKMKKSESGFNIP 425
QY 721 TMKEANFWANFYVNSGDCGHFTWDRHTAHLMLLEADNHVYVNCLOP 769
DB 426 LVI-----AGVFSGLVAGVYVFNKKKERENI.LTVDEKELTTIAP 467

RESULT 31
ID Q23587 PRELIMINARY; PRT; 3507 AA.
AC Q23587;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE ZK783.1 protein.
GN ZK783.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirschen J., Lister J., Latreille P.,
RA Lightfoot J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";

```

RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [3] SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U13646; AAC24418.1; -;
 DR HSPB: P00736; IAPQ.
 DR Interpro: IPR000152; Asx_hydroxyl.
 DR Interpro: IPR000561; EGF-like.
 DR Interpro: IPR001881; EGF Ca.
 DR Interpro: IPR001507; Endoglin/CD105.
 DR Interpro: IPR001254; Ser_protease_inh.
 DR Pfam: PF00008; EGF_15.
 DR SMART: SM00179; EGF_CA; 12.
 DR SMART: SM00001; EGF_16.
 DR SMART: SM00241; ZP_1.
 DR PROSITE: PS00022; ASX_HYDROXYL; 9.
 DR PROSITE: PS00186; EGF_2; 22.
 DR PROSITE: PS01187; EGF_CA; 13.
 DR PROSITE: PS00135; TRYPsin_SER; UNKNOWN_1.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SQ SEQUENCE 3507 AA; 365315 MW; 154F0B687874D9DF CRC64;

Query Match 4.48; Score 198; DB 5; Length 3507;
 Best Local Similarity 20.68; Pred. No. 0.00096;
 Matches 160; Conservative 112; Mismatches 282; Indels 222; Gaps 33;

QY 77 ISNPKRVLTTRSGRRANIFSAFLPCFNDKQIVSGSGGVFYNVQDADFTNRQCC 136
 DB 1994 ISSYTSKDMTSSKSPNVMSSSPVSSSSSTASSETV-----SIPSESSSSSEAP 2049
 QY 137 FTCHYGTYYIMVPPNDPYFLSCGEGDYRWFEDTRIKTSCTKEDCKDILINCRATS 196
 DB 2050 LTSSPATTEVIT-----ESSVSKTPKRESSSEITVKLSKSPE 2089
 QY 197 VALCPPIYLAAGCSDSVRIYDRMLGTRANGVAGRGTCGVAFIFSHLNKRCRY 256
 DB 2090 V-----TESSK-----SSSTPSTQSQSVTVPETSKSTVLSAPV 2128
 QY 257 TSL-----CYSBDQOEIIIVSYSDYIYLPDKDTPARELKTSPSAEERRE----- 300
 DB 2129 TSTSPTEVHTSSETPKSLASSTGTGDTNSTPTSSLASVKSASAPGTASAVPVKLS 2188
 QY 301 ---ELNQPVPKRL-----RLRGDMDSTGRAPRESERE-----RDEQSPNVSQ 343
 DB 2189 LSPDVQSPSTKTPDAEESTVQASSTSVKSTSEPSHVTKLITSSNPSSSVPTS 2248
 QY 344 RMSDMLSRWFEASVQAQNRGRGRS-RPRGCTSO-----SDISTLPTVP 387
 DB 2249 PKS---TPVPESTEQPTSTTPSGQSLTPMANSSEVLTTEPHVLSLSLPPDVQSSTTP 2305
 QY 388 S-----SPDLFVSTAMEVD-----TPAQPLQ-----STSTMSQAHS-- 423
 DB 2306 NNLSESVETPTKTSVSLNSESEPTTEAPTLSPDILSTTNLSQSTSTEDRSEI 2365
 QY 424 TSSPTSPHSTPL-----SSPD---SEGRQSVASGHHTHSDNNNEKL---SP 468
 DB 2366 SSNSKSTPAPLTVSSVTHVAASSSPDYTESSEPDLLGSSSTENPEASSKQTLISSTP 2425
 QY 469 KPCT---GEPLVSLHYSTEGTITSTI-KLINFTEWSS---TASSRSGISCKSRGCE 519
 DB 2426 TPDITTAASEPSTKSTMSPOLSTISNVLSSTTPPESSSKSPVSSSTEGISVYTSF---E 2482
 QY 520 ESTVPPSSVQ-----PPGDSSTKAPESSESDVTK-----YQGVSAEN- 558
 DB 2483 FSKVPESTTISVLEDLTKTTPBPILEETTTASETSEPLEDLTVSVRIHEILTSENV 2542
 QY 559 PVENHNINQSDKFTAKPLDLSNGERNDLNDKSGVPESSASSEKAKEE----- 609
 DB 2543 PRESESTTSSS--SSKPSOEPRAG-----ILSTVVPVTSVSLITASELTATISNTPPK 2595

QY 610 -----TSDQSTESATNENN-TNPEFOFQEATGPSA-----HEETSRDSAL--- 651
 DB 2596 QGRFTTSPKSLVASTSPSVTSSEPSSEKTRKTVSTTVSTTPTEETTESLILTA 2655
 QY 652 -----QPTDSDDDPVLIPGARYRAGGDRSAVARIOEFFRRKREKMEELDT 701
 DB 2656 APSKPTSTESSEAP-TPAKTSETPSNVST-----SKRSTENVT 2698

RESULT 32

ID Q9DGL1 PRELIMINARY; PRT; 791 AA.
 AC Q9DGL1;
 DT 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
 DE Retinitis pigmentosa GTPase regulator-like protein (Fragment).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neoteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_Taxid=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vervoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Milano M.G.,
 RA Melnd A., Mellinger T., Ciccociolla A., Wright A.F.,
 RT "Mutational hot spot within a new RGR exon in X-linked retinitis
 RT pigmentosa".
 RL Nat. Genet. 0:0-0(2000).
 DR EMBL: AF286475; AAC00554.1; -;
 FT NON_TER 1
 SQ SEQUENCE 791 AA; 89672 MW; DA40A0FA99E3E39 CRC64;

Query Match 4.38; Score 195.5; DB 13; Length 791;
 Best Local Similarity 19.98; Pred. No. 0.00019;
 Matches 122; Conservative 90; Mismatches 294; Indels 107; Gaps 17;

QY 163 DGIYRWFEDTRIKTSCTEDCKDILIN-----CRANTSVAICPPPIYLAAGC 211
 DB 231 EGSGRLE-----KSSSPEDSGSVLRAQSESEAHGAFSKNSAVINIQPPASSSEEGS 286
 QY 212 SDSVRIYDRMLGTRATGNAG-----KGTGVANFIPSHLNKSCRYTSLCYSDGQ 266
 DB 287 QDNSESV-----TANEEDEKESIQGTEEEEEEEDQKEOKTSADVSEEE 336
 QY 267 EIIIVSY-----SVIYLFDPKDPARELKTSPSAEEREELRQPPVRLRLRGD 315
 DB 337 ENTVQSSDKASDEKEDSDTLH---PEDSQSDVEDEBEGESKSEQEEEEESEEE 393
 QY 316 WSDTPAPARESERERDGEQSPNVSLQMSDMLSRWFE---ASEVAQSNRGRGRPR 372
 DB 394 RSTF---AESESEDEKKGERK-----ERDVEEBEGSEVEEESGSDHSEK 437
 QY 373 GGTQSQSD-----ISLPIYVSSPDLEFVSTAMEVTPAPQFQPTSTSTMSQAHA 422
 DB 438 KKASESEEEEBEEDBSLSEEBGDSKDAEESDAESETGEGAEQBSSTERKESNEE 497
 QY 423 STSPTESPHSTPLSPDSEORQSVASGHHTHSDNNNEKLSPKPGTGEPLVSLHYS 482
 DB 498 EQSSTEKEEENEGQSTKEEESNEEQSS--TEKESNEDEBS-----TEKEES 547
 QY 483 TEGTTTSTILNFTDQKSTASSSRGIGSKCSGQESFPVQSSVPPGDSSTKAPAE 542
 DB 548 NEEQSSSTEKEEESNEEQSSTEKEEESNEEESSEESNEEESDEEDQDEEETGEE 607
 QY 543 SSEDVTKYQGVSAENPVENHNINQSDKFTAKPLDLSNGERNDLNDKSGVPESSASS 602
 DB 608 DEEEEBEESQONQEEAEEDDETDVQEEEEEETKEKEDE-----EEETEK 657
 QY 603 EKAKEPESDQSTESATNENNTPPEFOFQEATGPSAHEETSTRSALQDQDSDDDV 662
 DB 658 EEDDEEETDRVABEEDAENNEEESSEESDEEDSDGEEDEEDSDSEDEEESSE 717

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Aamathides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Adair J.F., Agayari A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew J., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Flooker A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jaitli M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reihert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Sprelding A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL: AE003765; AAF56794.2; -;
 DR EMBL: AE003765; AAF56795.2; -;
 DR HSSP: P12111; 2KNT.
 DR Flybase: FBgn0003137; Ppn.
 DR InterPro: IPR000361; EGF-like.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR002223; Kunitz_BPTI.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR002221; WAP.
 DR Pfam: PF00047; Ig_3.
 DR Pfam: PF00014; Kunitz_BPTI; 12.
 DR Pfam: PF00090; tsp_1; 5.
 DR Pfam: PF00095; WAP; 1.
 DR PRINTS: PR00759; BASICPTASE.
 DR ProDom: PD000222; Kunitz_BPTI; 12.
 DR SMART: SM00408; IgC2; 3.
 DR SMART: SM00131; KU; 12.
 DR SMART: SM00209; TSP1; 7.
 DR SMART: SM00217; WAP; 1.
 DR PROSITE: PS00317; 4 DISULFIDE CORE; 1.
 DR PROSITE: PS00280; BPTI_KUNITZ_1; 11.
 DR PROSITE: PS00279; BPTI_KUNITZ_2; 12.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS00092; TSP1; 3.
 KW Alternative splicing; Immunoglobulin domain;
 KW Serine protease inhibitor.
 FT VARSPLIC 2803 2803 L -> SVSVV (IN SHORT ISOFORM).
 FT VARSPLIC 2844 2854 ENFKTWDSDGI -> VASPLPHNAV (IN SHORT
 FT VARSPLIC 2855 3060 ISOFORM).
 FT VARSPLIC 3060 331579 MISSING (IN SHORT ISOFORM).
 SO SEQUENCE 3060 AA; 331579 MM; ACA31D3E558C7C0 CRC64;

Query Match 4.3%; Score 194.5; DB 5; Length 3060;
 Best Local Similarity 20.4%; Pred. No. 0.0014;
 Matches 96; Conservative 77; Mismatches 171; Indels 127; Gaps 18;
 QY 259 LCYSEDSQGLVSVSSYITLFPKDDTA---RELKPSAEERRELRQPVKRLRGD 315
 DB 562 ICEPFDKTV-----PPADDCKCNKRETKESDDEGEERKVP-----GE 602
 QY 316 WSTGTP-----RARPESER----- 329
 DB 603 WF-TGPFGRKSKRCGGGERVCLSNGRKSYNCDEKEPESKNSACTEDELPL 661
 QY 330 -----ERDEGSPNVLQRMQMDLSRWFEEASVQAQNRGRSPRGTSQSDISTL 383
 DB 662 TSTDKPIEDDEDCDEDEGILISGLSD--DEKSEVIDIEGTAKE----- 706
 QY 384 PTYPSSDLEVSTFAMVDPAPQFLQPSSTMSNAHSTSPTEPHSTPLSSPDE 443
 DB 707 -TPPEADLMQSDS---PPEYDF--ESTGTPEGSGYDSESTPDSISTE--GSGDDE 757
 QY 444 Q-----RQSYEASGHTHHQSDNNNEKLSPKPGEPVLSTHYEGTTSTKLNF 495
 DB 758 ETSEASDLSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSST 812
 QY 496 TDWSSSIASSSRGIGSHCKSGEGESFVPOSSVQPPRGDSETKAPESSEDVTKYOGVS 555
 DB 813 -ETGVSDDSTVSSSTESASSESTVSGASDSTGTSNADSTPESTASSSTDDSTD 870
 QY 556 AENPVENHITQSDKRTAPLDSNGERDLNDRCGPESASSEKKEPETSQGT- 614
 DB 871 SS---DNSSVSSSTESASSVSDSSDSDGSTD--GVSTTENSDDSTSDATSDSTA 924
 QY 615 ---STESATENNTPPEOPOTATGPSAHESTRD--SALDPTDSDDD 660
 DB 925 SSDSTDS-TSDQTTETPESTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 974
 RESULT 35
 ID 090A02 PRELIMINARY; PRT; 1680 AA.
 AC 090A02;
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE Hypothetical 184.6 kDa protein.
 GN H24G06.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Rhabditidae; Rhabditidae; Rhabditidae;
 OC NCB1_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None.
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RL investigating biology. The *C. elegans* Sequencing Consortium.";
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Ryan E., Delenauty A.;
 RT "The sequence of *C. elegans* cosmid H24G06.";
 RN [3]
 RP Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF101309; AL32229.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 10.

DR SMART: SM00320; WD40; 11.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 DR PROSITE: PS0082; WD_REPEATS_2; 2.
 DR PROSITE: PS0294; WD_REPEATS_REGION; 2.
 DR Hypothetical protein.
 SQ SEQUENCE 1680 AA; 184618 MW; 856105233DE0FDA CRC64;

Query Match 4.3%; Score 194; DB 5; Length 1680;
 Best Local Similarity 22.1%; Pred. No. 0.00065;
 Matches 203; Conservative 119; Mismatches 402; Indels 194; Gaps 43;

QY 47 LNVHDCVNTICMNDICE-----YIISGSDTKLVIN--PYSRK-VLTTRSGHRA 95
 DB 512 INAHHDVYIDSDHASTSSHDPIPLASGGRDFYHARRRPIYSGQFHCAYLDHOS 571
 QY 96 NISAFELPCTNDKQI--VCSGSDGYFTYNEQDAETNRQOFT---CHYGTIEM 148
 DB 572 AIKSIK--ASNNGQLHLYTAASDRSLIIV--KLISFSDHCHFTYQIMSVASSIGDM 626
 QY 149 TVPNDPTFLSCGDEGIVRFDRIRKTSCKECKDILINCRANISVAICPIPYLA 208
 DB 627 NFKHVDLFYVACHDRMLRQFDINGKTVREVKGTDDVAHSGK--ILKVALDHSGSYAIS 684
 QY 209 VCGSDSVRIYDRMLCTRATGNVAGRTGMVARFIPSHLNKNSCHVTSICVSEGOEI 268
 DB 685 V-CSDKFTYITDLR-----SGVCL---AVLCGGAATDATTSDDFKNV 724
 QY 269 LVSYSSDYIYLFDP-PKDDTARELKTPSAER-REELRQPPYKRLRLGDWS---DTGP 321
 DB 725 IYVTSNCSIFIMOLAKMLTERMI--SAQVRLMEVMTATPDSLIGSGSETISGDSNS 781
 QY 322 RAPESEEREDGROSPVVISLMQRMSDMLSRWFEEASVAOSNRGRSGSRPGG----- 374
 DB 782 FGRLPGLAPESSG--SASLYISDDDS---TRFSSSVRSRK-ILPGLGLIDSS 832
 QY 375 -TQSQDSTPTVPSPSPDLR-----VSETAMEVDTPAEQFLQPSSTSSQAHSSTSS 426
 DB 833 YAVGDSSEFPAYQSAVAVERRTINLFSDQYETDVSFQ--SDYSSRRKTKLFDSD 890
 QY 427 PTESPHSTPLISPDSEBOROSVEASGHHTH---HOSDNNEKLSPKPGTG----- 473
 DB 891 QDSNLGSAQVLAAPINEDRRSASPILYRPODLRGYQSSSMNMLRVYTGAGVAVQAKE 950
 QY 474 ---EPVSLIHYSTEGTITIKLNFDP--EWSSTIASSSRIGSHCKS--EEOGESFPVQS 526
 DB 951 LMSQILASQRNSQSSGSHLSSTNTSGRMWGMDPQOSNNNDWHPSTVDIHVSPWAT 1010
 QY 527 SVQPE-----GDSETKAPEESSED---VTKYQGVSAENVEENHI-NITQSDKPTA 574
 DB 1011 SMIPPOHOGYGRDHTSTFMPDSTHPPPLAPRTSRVLSTASQALQOIOASSPFR 1070
 QY 575 KPLDSNGERNDL-----NLDRCGVPEESASSEKAKEPETSDQSTESATN 621
 DB 1071 KSMDRNLSRRLFLNGGAQOPTVWSPALANOGAPRRSNSNLFATNLLEVPTSTN--LS 1128
 QY 622 ENNTNPPQ---IQTETGSAHEHETSTRSALOD-TDDSDDDVYLIP--GAKRAGAGD 675
 DB 1129 RRHTDKOPKLRFTT-----TVQTRISLDEVEVSDDAITSTTFPEKKNRIGY 1179
 QY 676 RRSVAARIQ---EPPRRR--KERKMEELDTLIRPL-----VKWYKKGHRNSR- 720
 DB 1180 RRSSTVVDTRDAVARRAVAGTSPKSDRADDLNPAASSLSRSGSPKFLSOMINRD 1239
 QY 721 TMKEANFWGANFY-----MSGSDCGHIFTMDRHTAEHMLLEADNHVYVNCLOPHFD 773
 DB 1240 TPSSPSSSTGSIYGRMLQORRSGSVSYVSRATRGREMKRSDALNKL----- 1289
 QY 774 PILASSGIDVDIKITWSPLEBSRIRNKLIADLV-ITRNEIMLEIRNTTIVASFTMLRLA 832
 DB 1290 -----MAVRSKHOSSENILRSTENLALIKN--LEEASNSTPRA---RTSR 1331
 QY 833 SLNHIRADRLGDRSGS 850
 | | | | |

DB 1332 SSNLRNADNLGNMESFCT 1349

RESULT 36
 ID Q9P785 PRELIMINARY; PRT; 451 AA.

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Hypothetical serine-rich repeat protein.

OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC NCBI_TaxID:4896;
 RN 11

RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Aert R., Volckaert G., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL353012; CAB88235.1; -

SQ SEQUENCE 451 AA; 47427 MW; 36F254388048FC27 CRC64;

Query Match 4.3%; Score 193; DB 3; Length 451;
 Best Local Similarity 22.8%; Pred. No. 0.00013;
 Matches 102; Conservative 63; Mismatches 170; Indels 112; Gaps 15;

QY 284 DDTARELKTPSAERREELRQ-----PYKRLRLRGDWGDTGPRAPESEREDGROS 336
 DB 64 EDVGRKHKTKSLKSNDSOKISKKGAPRK-----AHSSEASGSGSS 109
 QY 337 PNVSIMQMSDMLRWFEASEVAOSNGRGRSRRGTSQSDISTLPTVPSPDLEYSE 396
 DB 110 DE-----SDSSSESSSESDSSSDSSSDSSSESSSESDSS-----SSSDSE-SE 156
 QY 397 TAMEVDTPAEQFLQPSSTITSQAQHSSTPESHSTPLSPDSPEOROSVSAHGHTH 456
 DB 157 SSSSE-----GSSSSSSSSSESSSESDSSS---SSDSSESSSEDS-SSS 202
 QY 457 HQSDNNNEKLSPKPGTGPVLSLHYSTEGTITIKLNFDEWSSIASSGIGSHCKSE 516
 DB 203 SSSDSSESS-----SBRGSDSSSSSSSESSSESDSSSESDSSSESDSSSE 246
 QY 517 GQESFVPOSSVQPEQSETKAPEESSEDTYKQGVSAENPVENHINITQSDKFTAKP 576
 DB 247 SSSSESDSSSSSDSESSSKSDSSSDSSSDSSSDSSSDSSSDSSSDSSSDSS 305
 QY 577 LDSNGERNDLNLDRCGVPEESASSEKAKEPETSDQSTESATNENTNPPQOTAT 636
 DB 306 SSSSSSSSE-----DGNSTDTTSGEVSQSSSTNSTSEESTSVKDESSKIH 355
 QY 637 GPS---AHEE---TSTRDS-----ALQPTDSDDDVYLIPGAR 668
 DB 356 DKSIAKRKHEDDESSSTSTSTTKPTTRVGDPSQWDAPSRALRNSNPFEDDY----- 409
 QY 669 YRAGGDRSAVARIQEFFFFRRKKE 695
 DB 410 ---GTLANRDLIVTGKGFQEKKKK 433
 | | | | |

RESULT 37

Q9S255 PRELIMINARY; PRT; 852 AA.

AC Q9S255;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-OCT-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE Hypothetical 92.0 kDa protein.
 GN F1IC18.80 OR A14G3180.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
 RA Villareel R., Glielen J., Van Montagu M., Homelisel J., Mewes H.W.,
 RA Mayer K.F.X., Scheller C.;
 RL Submitted (Apr-1999) to the EMBL/genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (Apr-1999) to the EMBL/genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
 RA De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villareel R.,
 RA Glielen J., Van Montagu M., Mewes H.W., Lemcke R., Mayer K.F.X.;
 RL Submitted (Mar-2000) to the EMBL/genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (Mar-2000) to the EMBL/genbank/DBJ databases.
 DR EMBL; AL049607; CAB40758.1; -
 DR EMBL; AL161579; CAB79906.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 852 AA; 91950 MW; AEK128F4D5A0F59D CRC64;

	Query Match	Best Local Similarity	Score	Length
			4.3%;	852;
			21.9%;	Pred. No. 0.0003;
	Matches 129;	Conservative 76;	Mismatches 224;	Indels 160;
			Gaps 26;	
OY	162	EAGTWRMPDTRTKTSCTKEDCKDDILINCRATSVATCIPPIYLAVGSCSSVRITDR	221	
Db	365	ESSSQKQADLS-KSDDIKEETEPABELDSKVLTSPV-----DSSV-----	405	
OY	222	RMLGSTRATGNVAGRGTGMVARFISHLNKSRCRVTSLCYSDEGGELLVYSDDIYLFDP	281	
Db	406	---TAATSS-----ENENKNKV-----QLPSKTSIG-----	428	
OY	282	PKDDTARELKTPTSAERREELNQPYVKRLRLRGWSDTGPRARPEREREDEQSPPNSL	341	
Db	429	--DETA-NVSSFSMAEELPEQSEVP-----KTLANOKKKESSTE--EVKPSASI	471	
OY	342	MQRMSDMLSRRPEEAASEVAOSNRGRSRPBGSTSGSODISTLPYP-----SSPDLEV---	394	
Db	472	AT-----EEVSEEPNTSEPO-VTKSKSKYAASSKTIKPYTPPSKSTSETKVAKO	520	
OY	395	-----SETAMENDVPAEQFLQPSTSTSMQA-HSTSSPRESP--HSTPLISSPDSEO	444	
Db	521	SEKRIVGSDNQMEQSPRKPEEKKKPGKGALIDEESLHTSSGDNEKRAVSGKLASKSKKA	580	
OY	445	KQSVASGHHTHHGSDNNNEKLSPPRGTEPVLSLIHSTEGTTSTIKINTDEMSSIAS	504	
Db	581	KQTVF-----ESPNSNTKRKRSLLGGOKA-----SGESELVGSRIRV---WWPMQA	622	
OY	505	SSRGIQSHCKS-----EGQEESRPVQSSVOPPEDSDSTKAPDES-----EDVT	548	
Db	623	YTKGVVESTDAKKKHLHYTDGDEELIYLKMQKNSPLDESELSDODEAAQDTGOEEPAS	682	
OY	549	KYQEGVSAE--NPVENHINITQSDKFTAKPLDENSEGERNDLNDRCGVGEESASSEK	604	
Db	683	TYGSGAGSGSKAKATPASKSSKTSQDDKTAKSKSDCKEASR-----EEBASSEE	730	
OY	605	AKEPTSPQT---STESATNENNNTNPEFOFOTEGATGPNAHEETSTRDSALOPTDSDDP	661	
Db	731	ESEEEEPRTYVKGSGSSSKRDISSVSKSGSKASSKKKEEPSKATTS----KSKGP	785	
OY	662	VLIPEARLRAGGDRSARVANRIQEFFRRRKREKKEHELDLTINRPLVK	710	
Db	786	VKSPVAKKGTGKGAKSGSASTPA-SKAKESAEESESEETPEKPPARK	833	

DB 677 TAOSTPLASONPT-----SSGTGTHNTPRTYPTPHQKIXENTENKLSFPTVASE 729
 QY 696 MEDLDLNTLRRL--VKANYKGRN-----SRVTK-----EANFGANEV-----MS 736
 DB 730 FHEMSTAESOTPLDVKYIVEKFSNDGEVATCSTVKSPEYVETN-MKVDLDVDMDEIS 788
 QY 737 GSDCGHIF-----IMDRHFAEHLMLLEADNHVYNLOLQHP--FPDILASSG 780
 DB 789 GNSPAGVENSNEKOKOLYRYVDGRISYQMLCSTSH-----SPEPCYCLFTSLIAR- 842
 QY 781 IDYIKIWSPLERSRIFNRKLADEVITRNLMEETRNITVTPASPMALASLNHI 837
 DB 843 -----EKODIAELKFTSD--PQTAVCTITLPSGVVPRREMLNNV 880
 RESULT 39
 Q9W320 PRELIMINARY; PRT; 2951 AA.
 AC Q9W320: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE CG3950 protein.
 GN CG3950.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydrioidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modarres G., Morris J., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.T.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relvert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Weissbach J., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003438; AAF46171.1; -
 DR Flybase; Fggn0029875; CG3950.

SO SEQUENCE 2951 AA; 332269 MM; 85CCOABBDAC8AEB CRC64;
 Query Match 4.2%; Score 191.5; DB 5; Length 2951;
 Best Local Similarity 21.5%; Pred. No. 0.002;
 Matches 108; Conservative 72; Mismatches 181; Indels 141; Gaps 22;
 QY 273 SSDYIYLPDKRD-TAEELK-----TPSAEERREELROPVRLRLRCWSPTG-PRAPES 327
 DB 455 TSEIVEYEDRDRODPTSHVQAKAPPSPYAPPSHSPROSPAK-----DFSTHGFPSPVPRNK 508
 QY 328 EREDEGQSPNVS-----LMQRM-----DMLRWFEASVQAQNR-- 364
 DB 509 ATQETPEQRCIGAGEEVVYVRSEKSRQYKQOTSQRTIEVEGDDYQEPQRFQKIREA 568
 QY 365 ---GRGRSRPRGGTSQSDIST-LPTV-----PSSPDLEVSETAMEYDTPAEPLQAP 411
 DB 569 PTPSWEDPQPRQPVVEEDFTHGFPVRYTTTSTRPQDGEVLHTSKTVSRNQSARKT 628
 QY 412 STSTMSAQAHSTSSPTESHSTPLSSPPSQSQSVASGHHTHOSDNNEKLSRPG 471
 DB 629 NTERLIETQVHPNAPS--HSTPRSGSPRRSQPRDYAS-----SPRGP 670
 QY 472 TGEPLVSLHSTEGTSTIKLNTDE-----WSSISS 506
 DB 671 AGRPQSQRPNPTGSGITTTTSEPLIRRLQKREVDAAHRAAASLRSSSPADST 728
 QY 507 RGISGHSCKSEGOSEFVQSSV-----OPRGDSETRAPESSE---DYTKYQEG- 553
 DB 729 TSYGSHHQ-----PRSSVSNPTFRREMREGSHDSQAPSSSSISSTYTRHHTGG 780
 QY 554 -----VSAENPVENHNITQSDKFT--AKPLDSNGERNDNL--DRSCGVPEESASS 602
 DB 781 NYSNTIKTKITTRPNHKKKAIEPATDTSQPIRKVILSANEKAVVEACVRRQYOLG 840
 QY 603 EKAKEPETSQTSSTESATNEN-----NTNPEQFQTEATGQSAHEESTDSAL----- 651
 DB 841 NGENENPTPSSSGTPAKKRQHOMRRPDEPPEQLRRSSKSPSVPRQVQKETTPEGRRVS 900
 QY 652 QDTRDDDDDPVLT-----PGA 667
 DB 901 QDREISIDELLILEETSGAGS 922
 RESULT 40
 Q9R0L0 PRELIMINARY; PRT; 1508 AA.
 AC Q9R0L0: 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE JNK-binding protein JNKBP1.
 GN JNKBP1 OR JNKBP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRN;
 RX MEDLINE=99402957; PubMed=10471813;
 RA Koyano S., Ito M., Takamatsu N., Shiba T., Yamamoto K., Yoshiooka K.,
 RT activation of JNK by MEK kinase (JNK)-binding protein that enhances the
 RL FEBS Lett. 457:385-388(1999).
 CC -1- SIMILARITY: CONTAINS 12 WD REPEATS (TRP-ASP DOMAINS).
 DR EMBL; AB029482; BAA85449.1; -
 DR MGI; MGI:1347004; Jnkbp1.
 DR InterPro; IPR002114; HPr_Serp_site.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 12.
 DR SMART; SM00320; WD40; 11.
 DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
 DR PROSITE; PS50082; WD_REPEATS_2; 1.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:35:56 ; Search time 17 Seconds

(without alignments)
2331.717 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRGSYPHLLWDVKKRSLG.....RLEGDRSGQENEDDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	569	12.6	110	10	US-09-864-761-35853
2	437	9.7	86	10	US-09-864-761-35853
3	338	7.5	62	10	US-09-864-761-33782
4	273	6.0	214	10	US-09-925-300-1585
5	190.5	4.2	1367	10	US-09-801-368-108
6	189	4.2	1142	9	US-10-085-108-7
7	189	4.2	1142	10	US-09-899-651-2
8	188	4.2	1702	9	US-09-839-996-5
9	185.5	4.1	571	10	US-09-765-272-4
10	183	4.0	2344	10	US-09-815-242-12713
11	180.5	4.0	1175	10	US-09-771-161A-224
12	180.5	4.0	1175	10	US-09-771-161A-225
13	180.5	4.0	1175	10	US-09-771-161A-226
14	179.5	4.0	1501	10	US-09-924-154-17
15	178	3.9	1325	10	US-09-864-761-35612
16	175.5	3.9	1848	9	US-09-839-996-6
17	174	3.8	2368	10	US-09-815-242-5635
18	174	3.8	2368	10	US-09-815-242-12389
19	170	3.8	684	10	US-09-925-302-555

120	169	3.7	1421	10	US-09-924-154-13	Sequence 13, Appl
21	167.5	3.7	1031	10	US-09-815-242-10932	Sequence 10932, A
22	166	3.7	560	10	US-09-815-242-13057	Sequence 13057, A
23	165.5	3.7	1435	9	US-10-153-273-4	Sequence 4, Appli
24	164.5	3.6	957	9	US-10-025-380-1065	Sequence 1065, Ap
25	164.5	3.6	957	10	US-09-922-217-1065	Sequence 1065, Ap
26	164.5	3.6	957	10	US-09-833-263-1065	Sequence 1065, Ap
27	164.5	3.6	2478	10	US-09-815-242-5816	Sequence 5816, Ap
28	164.5	3.6	2478	10	US-09-815-242-12967	Sequence 12967, A
29	164	3.6	278	10	US-09-765-272-94	Sequence 94, Appl
30	160.5	3.5	1356	9	US-10-077-111-10	Sequence 10, Appl
31	160	3.5	502	10	US-09-815-242-5904	Sequence 5904, Ap
32	156.5	3.5	713	10	US-09-801-368-408	Sequence 408, Ap
33	155.5	3.4	505	9	US-09-820-843A-12	Sequence 12, Appl
34	155.5	3.4	505	10	US-09-815-242-11317	Sequence 11317, A
35	155.5	3.4	617	10	US-09-864-761-36182	Sequence 36182, A
36	154.5	3.4	729	9	US-10-145-396-11	Sequence 11, Appl
37	152.5	3.4	980	9	US-10-108-605-195	Sequence 195, Ap
38	152.5	3.4	1781	10	US-09-738-877-3	Sequence 3, Appli
39	151.5	3.3	1139	9	US-09-820-843A-15	Sequence 15, Appl
40	151	3.3	1075	10	US-09-801-368-110	Sequence 110, Ap
41	149.5	3.3	748	10	US-09-864-761-43244	Sequence 43244, A
42	149.5	3.3	1322	10	US-09-801-368-114	Sequence 114, Ap
43	149	3.3	473	10	US-09-864-761-38321	Sequence 38321, A
44	149	3.3	1162	10	US-09-745-008-34	Sequence 34, Appl
45	148.5	3.3	897	10	US-09-815-242-12769	Sequence 12769, A
46	148.5	3.3	1113	10	US-09-815-242-5836	Sequence 5836, Ap
47	148.5	3.3	1596	9	US-09-902-432-4	Sequence 4, Appli
48	148	3.3	1125	9	US-09-974-298-114	Sequence 114, Ap
49	148	3.3	1163	9	US-09-932-257A-3	Sequence 3, Appli
50	147.5	3.3	511	10	US-09-864-761-34590	Sequence 34590, A
51	147.5	3.3	2843	9	US-09-987-482-1	Sequence 1, Appli
52	147	3.2	1360	10	US-09-871-916-2	Sequence 2, Appli
53	146	3.2	1537	10	US-09-801-368-104	Sequence 104, Ap
54	145.5	3.2	596	9	US-10-063-547-100	Sequence 100, Ap
55	145.5	3.2	596	9	US-10-174-590-310	Sequence 310, Ap
56	145.5	3.2	596	9	US-10-176-758-310	Sequence 310, Ap
57	145.5	3.2	596	9	US-10-063-616-100	Sequence 100, Ap
58	145.5	3.2	596	9	US-10-175-737-310	Sequence 310, Ap
59	145.5	3.2	596	9	US-10-063-502-100	Sequence 100, Ap
60	145.5	3.2	596	9	US-10-173-706-310	Sequence 310, Ap
61	145.5	3.2	596	9	US-10-175-738-310	Sequence 310, Ap
62	145.5	3.2	596	9	US-10-175-752-310	Sequence 310, Ap
63	145.5	3.2	596	9	US-10-176-482-310	Sequence 310, Ap
64	145.5	3.2	596	9	US-10-176-913-310	Sequence 310, Ap
65	145.5	3.2	596	9	US-10-176-913-310	Sequence 310, Ap
66	145.5	3.2	596	9	US-10-180-552-310	Sequence 310, Ap
67	145.5	3.2	596	9	US-10-180-557-310	Sequence 310, Ap
68	145.5	3.2	596	9	US-10-173-700-310	Sequence 310, Ap
69	145.5	3.2	596	9	US-10-174-572-310	Sequence 310, Ap
70	145.5	3.2	596	9	US-10-174-579-310	Sequence 310, Ap
71	145.5	3.2	596	9	US-10-174-582-310	Sequence 310, Ap
72	145.5	3.2	596	9	US-10-174-588-310	Sequence 310, Ap
73	145.5	3.2	596	9	US-10-175-739-310	Sequence 310, Ap
74	145.5	3.2	596	9	US-10-175-740-310	Sequence 310, Ap
75	145.5	3.2	596	9	US-10-175-743-310	Sequence 310, Ap
76	145.5	3.2	596	9	US-10-176-488-310	Sequence 310, Ap
77	145.5	3.2	596	9	US-10-176-493-310	Sequence 310, Ap
78	145.5	3.2	596	9	US-10-176-747-310	Sequence 310, Ap
79	145.5	3.2	596	9	US-10-176-750-310	Sequence 310, Ap
80	145.5	3.2	596	9	US-10-176-985-310	Sequence 310, Ap
81	145.5	3.2	596	9	US-10-176-987-310	Sequence 310, Ap
82	145.5	3.2	596	9	US-10-176-991-310	Sequence 310, Ap
83	145.5	3.2	596	9	US-10-176-992-310	Sequence 310, Ap
84	145.5	3.2	596	9	US-10-176-993-310	Sequence 310, Ap
85	145.5	3.2	596	9	US-10-184-658-310	Sequence 310, Ap
86	145.5	3.2	596	9	US-10-173-695-310	Sequence 310, Ap
87	145.5	3.2	596	9	US-10-173-697-310	Sequence 310, Ap
88	145.5	3.2	596	9	US-10-173-705-310	Sequence 310, Ap
89	145.5	3.2	596	9	US-10-174-576-310	Sequence 310, Ap
90	145.5	3.2	596	9	US-10-174-585-310	Sequence 310, Ap
91	145.5	3.2	596	9	US-10-174-586-310	Sequence 310, Ap
92	145.5	3.2	596	9	US-10-175-747-310	Sequence 310, Ap

93 145.5 3.2 596 9 US-10-176-481-310 Sequence 310, App
94 145.5 3.2 596 9 US-10-176-485-310 Sequence 310, App
95 145.5 3.2 596 9 US-10-176-487-310 Sequence 310, App
96 145.5 3.2 596 9 US-10-176-493-310 Sequence 310, App
97 145.5 3.2 596 9 US-10-176-756-310 Sequence 310, App
98 145.5 3.2 596 9 US-10-176-911-310 Sequence 310, App
99 145.5 3.2 596 9 US-10-176-919-310 Sequence 310, App
100 145.5 3.2 596 9 US-10-176-925-310 Sequence 310, App
101 145.5 3.2 596 9 US-10-176-978-310 Sequence 310, App
102 145.5 3.2 596 9 US-10-179-510-310 Sequence 310, App
103 145.5 3.2 596 9 US-10-180-543-310 Sequence 310, App
104 145.5 3.2 596 9 US-10-180-544-310 Sequence 310, App
105 145.5 3.2 596 9 US-10-180-546-310 Sequence 310, App
106 145.5 3.2 596 9 US-10-180-547-310 Sequence 310, App
107 145.5 3.2 596 9 US-10-180-549-310 Sequence 310, App
108 145.5 3.2 596 9 US-10-180-555-310 Sequence 310, App
109 145.5 3.2 596 9 US-10-180-559-310 Sequence 310, App
110 145.5 3.2 596 9 US-10-181-000-310 Sequence 310, App
111 145.5 3.2 596 9 US-10-183-010-310 Sequence 310, App
112 145.5 3.2 596 9 US-10-183-012-310 Sequence 310, App
113 145.5 3.2 596 9 US-10-184-614-310 Sequence 310, App
114 145.5 3.2 596 9 US-10-184-623-310 Sequence 310, App
115 145.5 3.2 596 9 US-10-184-635-310 Sequence 310, App
116 145.5 3.2 596 9 US-10-184-637-310 Sequence 310, App
117 145.5 3.2 596 9 US-10-184-646-310 Sequence 310, App
118 145.5 3.2 596 9 US-10-184-647-310 Sequence 310, App
119 145.5 3.2 596 9 US-10-184-652-310 Sequence 310, App
120 145.5 3.2 596 9 US-10-187-594-310 Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-864-761-35853
; Sequence 35853, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35853
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL033531.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BT47A, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT47A, SIGNAL = 1.9
; OTHER INFORMATION: EST HUMAN HIT: A1927915.1, EVALUATE 8.00e-35
US-09-864-761-35853

Query Match 12.6%; Score 569; DB 10; Length 110;
Best Local Similarity 100.0%; Pred. No. 7.2e-28;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGERNDLNRSCGVPESASSEKAKEPETSQDTSTESATNNTNP 627
Db 1 QSDKFTAKPLDLSNGERNDLNRSCGVPESASSEKAKEPETSQDTSTESATNNTNP 60
QY 628 EPQFQTEATGSAHEETSTDSALQDITDDSDDDPVLIPGARYAGPGR 676
Db 61 EPQFQTEATGSAHEETSTDSALQDITDDSDDDPVLIPGARYAGPGR 109

RESULT 2

US-09-864-761-33782
; Sequence 33782, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 33782
LENGTH: 86
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031287.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81
OTHER INFORMATION: EST_HUMAN HIT: 238367.1, EVALUE 5.00e-25
OTHER INFORMATION: SWISSPROT HIT: Q64752, EVALUE 3.10e+00
OTHER INFORMATION: EST_HUMAN HIT: BE867206.1, EVALUE 5.00e-29
US-09-864-761-33782

Query Match 9.7%; Score 437; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 5e-20;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 GTSQSDISTLPTVPSPDLEVSSTAMEVDTAPQFLQPTSTSTMSQAHSSTSPHS 433
|||||
Db 1 GTSQSDISTLPTVPSPDLEVSSTAMEVDTAPQFLQPTSTSTMSQAHSSTSPHS 60
QY 434 TPLSSPDSFQSQVSEAGHHHQHS 459
|||||
Db 61 TPLSSPDSFQSQVSEAGHHHQHS 86

RESULT 3
US-09-864-761-33783
Sequence 33783, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-x-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 33783
LENGTH: 62
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031287.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EST_HUMAN HIT: BE019684.1, EVALUE 5.00e-32
OTHER INFORMATION: EST_HUMAN HIT: BE180609.1, EVALUE 8.00e-32
OTHER INFORMATION: SWISSPROT HIT: P34311, EVALUE 9.90e+00
US-09-864-761-33783

Query Match 7.5%; Score 338; DB 10; Length 62;
Best Local Similarity 100.0%; Pred. No. 3.1e-14;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 85 VLTIRSGHRANFSKFLPCTNDKQIVSCGSGDVIFVTNVEQDAETNRCQFTCHYGT 144
|||||
Db 1 VLTIRSGHRANFSKFLPCTNDKQIVSCGSGDVIFVTNVEQDAETNRCQFTCHYGT 60
QY 145 YE 146
|||
Db 61 YE 62

RESULT 4
US-09-925-300-1585
Sequence 1585, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:

QY 683 -----IQEPRRKRKE-----MEELD-----TLNRRPLVWVYKGRHSR 720
Db 939 ISRTGYEPVIFRKAREFIELFGISREVDPPDDSYFVNTLDLTSEGLSDEGMSQNR 998
QY 721 TWIKEANFGANFVNSGDCGHIFTD-----RHVAHMLLEADNHVNC--LQPH-- 770
Db 999 LILILSI-----IFKGYIAEEVNDVLSGIGVRAGREHFAFGEPRELLTKVWVQEHYL 1054
QY 771 PPDPLASSGIDYDIKWIPLESIFRKNLADEVITRNEMLLETRNT--ITVPASF 826
Db 1055 EYREVPNSPPRYEF-LWGPRAHSEVIRKVV-----FLAMKNTVPIITPSSY 1103

RESULT 8
US-09-839-996-5
; Sequence 5, Application US/09839996
; Publication No. US20030009010A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RET/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match 4.2%; Score 188; DB 9; Length 1702;
Best Local Similarity 18.9%; Pred. No. 0.0028;
Matches 150; Conservative 120; Mismatches 311; Indels 212; Gaps 33;

QY 67 LSGSDDTKLVTNSPYSRKVLTIIRSGHRANLFSKFLPCNTDKQIVSCSGGVIFVYTNVE 126
Db 773 ITASDNARVHIG--YKAGDTCVRSYDTG-----YVTCITDK--LSDKALNSFNATVNS 822
QY 127 QDAETNRQCQFCH-----YGR-----IYEMTVNDPDTYFLSCGEDGTVRWFDTFRKTS 177
Db 823 GNVNLGNNANFVLKANFLGHISGTGNSQVRLTNSHHLTG-----DSNVNQLND----- 874
QY 178 TKEDCKDDILINCRRAATVSAICPPIP-----YVLAVGSDSSVRIYDRMLGTR 227
Db 875 -----KGHHLNAQNDANKVTYNTLTVNSLSGNGSFYLL-----TDLNKGQGVVYTKS 925

QY 228 ATGNY-----AGRTTGWVARFIPSHLNNKSCRVTSLCYSED---GQELLVSYSSDYI 277
Db 926 ATGNFTLOVADTKGETPKNELTLFDASNATRNRLNLSVLTGNTVDLGAWYKLRNVNGRY- 984
QY 278 YLFDPK-----DDTAR-----ELKTPSAEEREEELRO-----PPVKRLRLRGDWS 318
Db 985 DLYNPEVEKRNQVDTNTITPNIQADVPSVPSNNEELARVETVPVPPAPATPSETTET 1044
QY 319 TGRAPES-----ERERDG-----EOSNVSLMQRMDSMLSRWE----- 354
Db 1045 VAENSQOESTVTEKNOQDATTETTAQNGEVAEEAKPSVKANTQINEVAQSGSEETETOTTE 1104
QY 355 --EASEVAQSNRGRGRPRGTSQSDISTLPVPS-----SPLEVSETA 398
Db 1105 IKETAKVEKEEKAKVEKEAKVEKDEIQEAPQMASETSPKQAKPAPKEVSTDTKVEETQ 1164
QY 399 MEVDTPAEQFLQPTSTSTMSAQAHSTSPSPESPHSTPLSSPDSEORQSVASGHHTHHQ 458
Db 1165 VQA-----OPOTQSTTTVAEAET-----SPNSKPAEETQPSKTNAPVTPVVKNO 1211
QY 459 SDNNEKLSKPGTGPVLSLHYSTEGTTITIKLNTDEWSSIASSSRIGISHCKSEGO 518
Db 1212 TENT-----TDQP-----TEREKTAKVETEKTOEPQVQASQA-----SPKQEQ 1249
QY 519 EESFVPOSSVQPPEDSETKAPESSEDEVTKYQGVSAENPV-ENHIN-----ITQSDK 571
Db 1250 SETVQOQVAVLESENVPVNNAAEYQAOLOQTQTSATVSTKOPAPENSINTGSAITETAE 1309
QY 572 FTAKPDLNSNGERNDLNDRCGVPESASSEKAKEPETSDQTSATSENNTNNTPPQF 631
Db 1310 KSDRP-----QTETAATEDASQHKANTVADNSVANNSESESEPKSRR 1351
QY 632 QTEATGSAHEETSTRDSALQDTSDDDDVLIIPGARYRAGPDRRSARVARIQIEFFRRK 691
Db 1352 RRSISQP---QETSAEETTAASDTDET-----TIADNSKRSKP-NRRS-----RRS 1392
QY 692 ERKEMEELDTLNIIRPLVKVYKGRHSRVTIKKANFVNSGDCGHIF-----WDR 748
Db 1393 VRSE-----PTVNGSDRSRTVALRDLTSTNTNAVISDAMAKAQFVALNVGK 1438
QY 749 HTAEHMLLEADN 761
Db 1439 AVSQHISQLEMMN 1451

RESULT 9
US-09-765-272-4
; Sequence 4, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-765-272-4

Query Match
Best Local Similarity 19.3%; Score 185.5; DB 10; Length 571;
Matches 99; Conservative 80; Mismatches 198; Indels 137; Gaps 17;

QY 212 SSSSVRIYDRMLGRATGNYAGRTGTGMVARFIPSHLNKSCRVTSLCYSDGDEILVLS 271
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 SDSQQLAEHKNLTKKKEKISPKETKGVNTLNPODEVLSGOLNKPILYREETMETKD 169
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 272 YSS-----DYILFD-PKDDTARELKTSPSAERRELRQP 305
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 FOEIQENPDLAGVTRVQKQGLKQKVEIVRIFSNNKEVSRIVSTST-----TA 221

QY 306 PVKRLRLRGDWSDTGPRAPSERERDGEOS-----PNVSLMQMSDMLSRWFEEASEVA 360
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 PSPRIVEKGTQVTKQPEGVGEHKDQSGAIVEPAIQ-----PELPEAV 268

QY 361 QSNRGGRSRPRGGTSQSDISTLPTV-----PSSPDLEYSETA-----S 790
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 VSDKGEPEVQ-----TLPEAVTDKETEVPQSPDTVYSDKGEPEQVAPLPEY 318

QY 399 -----MEVDTPAEQFLOPSSITSSAQASHSTSTSPHSTPLSSPDSEORQSVASG 452
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 KNISQVQKPEPTVEKTKQGGPEKTEEV-----PVKPTPEVPNPNEGTTEGTQIEAE 371

QY 453 HHTH--HOSDNNNEKSPKPGTGPVLSLHYSTEGTTSTIKLNTDDEWSSIASRRGIG 510
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 NPVQPAESTINSEKVP-----DTSKNTGEVSSNSDSTSTSGEKNKPEH 418

QY 511 SHCKSEGQESFVPOSSVQPPGDSGETKAPESSEDEVTKYQGVSAENPVENHINITQSD 570
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 NDSKNENSEKT-VEEVVPPNNEGTVEGTSNQETKPEVQPAER-----TQIN 463

QY 571 KFTAKPLDSNGERDNLDRSCGVPEESASSEK---AKEPETSQTSATNENNTNP 627
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 --SGKIANENTGEVSKFSDSKPPV-EESNQPKNGTKPKPNSGNTTSEN-----QTEP 516

QY 628 EPQFQTEATGFSAAHETSTRDSALQDTSDDDP 661
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 EPSNGNSTEDVSTESNTSNGNEEKQENELDP 550

RESULT 10
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes

```

```

; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12713

Query Match
Best Local Similarity 4.0%; Score 183; DB 10; Length 2344;
Best Local Similarity 21.3%; Pred. No. 0.0086;
Matches 135; Conservative 85; Mismatches 237; Indels 178; Gaps 25;

QY 51 DCVNTICWNDGEYILSGSDDTKLVISPNYSRKVTTTIRSGHRANIFSAKFLPCNDKQ 110
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 710 DNTNITISGPTN-----IGTSTIIVTSDAGNKKTTTTFKYEIVTRNSMS-----D 755

QY 111 IVSCSGDGVIFTVNQDAETNRQCFCHYGTYYEIMTVPNDPYFLSCGEDGTVRWF 170
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 SVTSGS-----TQQSQSVSTSKADQSASTSTSGSMT-----S 790

QY 171 TRIKTSCTKEDKDDILNCRRAATSAVAICPPPIYLAAGCSDSVRIYDRMLGRATG 230
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 TGASTS-----KSTSVSL-----SDS-----VSASKSLSTSEN 819

QY 231 NYAGRTGTMV-ARFTPSHLNKNKSCRVTSLCYSEDQGEILVSYSDYIYFLPKDDTARE 289
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 820 SVSSTSTSLVNSVSSSMGSGVSKSTSL-----SDFI-----853

QY 290 LKTPSAEERRELRQPPVKRLRLRGDWS-----TGPRAPSERERDGEQSPNVSLMQ 343
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 854 -SNSSTKESVSTSTSDSLRTSLSDSVSMSTSGSLSKSQS---LSTSTSDASTSQ 909

QY 344 RMSDMLSRWFEEASEVAQNRRGRSRPRGGTSQSDISTLPTV-----PSSPDLEYSETA 398
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 910 SVSDSTSNISSTESLSES-----GSTSES-ISISNISNSVSASTSKLESQSTS 958

QY 399 MEVDT-----PAEQFLOPSTSTMS-----AQAHSTSSPTESPHSTPLSSPD-----441
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 ISLSTSDSKSMSTSELSSTSTSDSVSGSLSVAGSQSVSTSTSDSMSEMISDSMSTS 1018

QY 442 ----SEQQSVASGHHTHHQDNNNEKLSKPKGTGE---PVLSLHYSTEGTTSTIKLN 494
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1019 GSLAAASDMSVSSSMSTSGSTSELSSTSTSDSKSLSTSGSGSTSTS---1074

QY 495 FTDEWSSSIASSRGIGCHCKSEGQESFVQSVQPPGDSGETKAPESSEDEVTKYQEV 554
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 -TSTSSVRMSQSSTSGSMSTSQSDSTSTSTSTSDSKS-ASTASSESTIS---QSV 1129

QY 555 SAENPVENHINITQSKFTAKPLDSNSGERDNLNDRSCGVPEESASSEKAKEPEPESDQT 614
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 STS-----TSGSVSTSTSLSTSNSTSTSM-----SDSTSLSTSESDSTSDST 1173

QY 615 STESATNENNTNPQFQTEATGFSAAHETSTRDS 649
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1174 STSDSISEAISG-----SESTISLSESNSTSDS 1202

```


RESULT 13

US-09-771-161A-226
; Sequence 226, Application US/09771161A
; Patent No. US2002011081A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 226
; LENGTH: 1175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-226

Query Match 4.0%; Score 180.5; DB 10; Length 1175;
Best Local Similarity 20.7%; Pred. No. 0.0049;
Matches 133; Conservative 86; Mismatches 260; Indels 163; Gaps 28;
QY 283 KDTARELTPSAERERELRQPPVKRLRLRGDWSGTGPRARPESEERERDGEQSPNVSLM 342
DB 373 EQQRERARRQEQERQEQEKR-RL-----EELRRRKEEERRRAEEKKRY--- 424
QY 343 QMSDMLSWFEAS---EVAQSN-----RGRGRSPRG 374
DB 425 EREQEYIRQLLEEQRHLVQLQQLLQEQAMLLHRRHPHQSPQPPPPQEQERSKSFH 484
QY 375 TSQSDISTLP-----TVSPSPDLEVSETAMEVDTP-----AEQFLQPSTSTM 417
DB 485 APEKHAHYEPADRAREVEDFRKTNHSSPEAQSKQGRVLEPPVPSPRSSEFSNGSESVH 544
QY 418 SA-QAHSSTSTESHSPVLLSSPDEQSQVEASGHHTHQSDNN----- 463
DB 545 PALQPAEPQVPTVTSRSPVLSRRDS-----PLQSGQONSQAGQNSTSSIEPRLWLR 600
QY 464 -EKLSPKPTGEPVLSLHYSTEGTTRSTIKLFTDWSSTIASSRGIG-----SHCKSE 516
DB 601 VEKLYPRPGS-----SSGSS-----NSGSPGSHPGSQSGGERFRVRSKSE 646
QY 517 GQEEFVPOSSVOPPGDSETRAP---EESSEDVTKYOGVSAENPVENHINITQSDKFTA 574
DB 647 GSPSQRL-ENAVKKPDCKKEVFRPLKPAGEVDLTALAKELRAVEDVRPHKVTD---YSS 702
QY 575 KPLDSNGERNLNL-----DRSCGVPEE--SASSEKAKEPETSDQSTSTESAINENTNP 627
DB 703 SEEGSTDEEDDDVBEQEGADESTSGEDTRAASSLNSNGETE---SVKIMVHDDVES 759
QY 628 EPQFQTEATGPSAHEETSTRDSALQTDSDO-----DPVLI----- 664
DB 760 EPAMTPSKETLIVRTQSNASTLQKHSSSFTFPDPLQLQISPSGGTVTVSVGFSC 819
QY 665 PGARYRAGPD--RSAVARIOEFFFFRRRKEEMELDTINIRPLVKVYKGHNSRTM 722
DB 820 DGMREPAIRQDPTKRGSVVNPNTNR-----PQSDTPEIRK-----YKRFNSEIL 866
QY 723 IKEANFWGANFVMSGDCGHIFWDRHTAEHLMLLE--ADNHVNVNCLQHPFPDPIASSG 780
DB 867 C--AALWGNVLLV-GTESG-----LMLDRSGQGVYPLINRRRFOQMDVLEG 911
QY 781 IDYIKINSPLEESRIFNRKLADEIVITRNLMLLEETRNITTV 822
DB 912 LNVLTISGKKDLRVYYSWLNKLNILHNDPVEKKQGTWTV 953

RESULT 14

US-09-924-154-17
; Sequence 17, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1501
; TYPE: PRT
; ORGANISM: Mammalian
US-09-924-154-17

Query Match 4.0%; Score 179.5; DB 10; Length 1501;
Best Local Similarity 17.7%; Pred. No. 0.0078;
Matches 104; Conservative 94; Mismatches 171; Indels 219; Gaps 22;
QY 285 DTARELTPSAERERELRQPPVKRLRLRGDWSGTGPRARPESEERERDGEQ-----SP 337
DB 851 ETLOSQEQLEAFAQKQKQEEEPKK-----QEEPPKKQEEQKQEQKQEQEQQE 903
QY 338 NVSLMQMSDMLSRWFEEASEVA---QSNRGRGRSR-----PRGQTSQSDIS-- 381
DB 904 KQEEQQTQDOSQSGLDOSKVGVASEQNEISSGQBNKSSSEVVPVQETTSENGSSQD 963
QY 382 -----TLPTVPSSPDLEVSETAMEVDTPAEQF 408
DB 964 TKISTEPENSVVDRAATSMNLDPEKVNENNNNDTNTPEASLKDCKEYD-DAKKE 1022
QY 409 LQPTSTSTWSAQAHSTSPTE-----SPHSTPLSSPDSE----- 443
DB 1023 LQSTVSRIESNEQDVQSTPPTPTVEGKVGDKAEMLTSPHAT--DNSESEGLNPTDD 1079
QY 444 -----QROSVASGHHTHQSDNNNEK---LSPKPTGEPVLSLHYSTEGTSTIK 492
DB 1080 IKTTDGVVYKEQILGGESATETSKNLEKPKDVEPSHEISEPVS-----GTT----- 1128
QY 493 LNFTDWSSTIASSRGISGSHCKSEQGESEFVQSSVQPPGDSSETKA--PEESSEDVTK- 549
DB 1129 -----GKESELLKSKIETKGTDPKRSNQEDATDDVVEN 1164
QY 550 -YQEGVSAENPVENHINI-----TQSDKETAKPLDSNSGE 583
DB 1165 SRDDNNLSNSVDNSQSNVLNREDPIASETEVVSPEDESSRIITTEVPSTTVKPPDKRSE 1224
QY 584 R-----NDNLND-----RSCGVPEESASSEKAKEPTSDQSTESATNENNTNPEQFQT 633
DB 1225 EVGEKAEKIEPVVYVPAIGPEMENSVS--VQSPNVNVEDVEKETLISENN----- 1273
QY 634 EATGPSAHEETSTRDSALQD-----TDDSDDDPVLIPGARYRAGPGDRRS 678
DB 1274 -----GLHNDTHRGNISSEKLDIHLNRNEAGSTLDDRRNGEMTEGS-----ES 1319
QY 679 AVARIOE--FFRRRRKERMEELDTINIRPLVKVYKGHNSRTMIK 724
DB 1320 DVGELEHNFSTQKDEKDFDQIASDREKEEIQKLLNIGHEEDEDVLK 1367

RESULT 15

US-09-864-761-35612
; Sequence 35612, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aecmca-x-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 35612
 LENGTH: 1325
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000511.1
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
 OTHER INFORMATION: EXPRESSED IN FEL LIVER, SIGNAL = 0.98
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
 OTHER INFORMATION: EST_HUMAN HT: AW867076.1, EVALUATE 1.00e-103
 OTHER INFORMATION: EST_HUMAN HT: AW867076.1, EVALUATE 1.00e-103
 OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATE 5.00e-63
 US-09-864-761-35612

Query Match 3.9%; Score 178; DB 10; Length 1325;
 Best Local Similarity 20.7%; Pred. No. 0.0082;
 Matches 144; Conservative 85; Mismatches 314; Indels 152; Gaps 30;

QY 59 WNDTGEYLSCSD-----DTKLVTNPNYRKVLTTIRSGHRANITSAKFLPCTNDKQTVS 113
 Db 7 YOKTGE--LSTSDHIFLTPGLVNSIPDPDHVL---HSGQR-----PPELPKSTEIH--- 53

QY 114 CSGDGVIFVYTNVEQDAENNRQCFT-----CHYGTYYEIMVDPNDPYTF 157
 Db 54 -----EOKRHCHNTFRHSKPTDKTCNSKTIDHKSSTDNHEAPPTSEENS 97
 QY 158 LSCGEGTV---RWFDTRIKTSCTEDC-KDDIL-----INCRRAATVAICPPTPYVL 207
 Db 98 SNOGKDPMIRNORSVDPADSTTHKESACKKHITAPKSKINCRAKSTTKS-----T 149
 QY 208 AVGCSDSVRIYDRMLGTRATGNAGRTTGWARFIPSHLNKNSCRVTSLCYSDGGE 267
 Db 150 VTRKSDKTGRPLEKSMISLDKTSSTSHKTTTSF-----HNSGNSQTKOKSTSFPE---K 200
 QY 268 ILVSYSDIYLYLFDPKDDTARE-----LKTPSAERREELRQPPVK 308
 Db 201 ITAASKTYKTTGTPESEKTEDSRTTVASDKLLTKTKNIQETISANELTQSLAEPT 260
 QY 309 RLRLRGDMSDTCPRARPERERDEQSPNYSLSMQMSDMLSRWFEEASEVAQSNRGR 368
 Db 261 GGTANE-NNTTSPAPTEPNERNTANENTTSPAEPT-----ENRERTANENTAPP 311
 QY 369 SRPRGGTSQSDISLTPTVPSPDLVSEVETAMEVDTPAEQLOPSTSTMSAQAHSTSS-- 426
 Db 312 AGPTENRENTANENTTLFPAETHEGERTANENTTSP--AEPTHEGERTANENTTSPA 369
 QY 427 -PTESPHSTPLSSPDSEOROSVEASGHHTHQSONNNEKLSPKPGTGEFVLSLHYSTEG 485
 Db 370 EPTHEGERTPFAN--DKTTSSAESTEH--GERTPLANENTTSP--AEPTENRERT 423
 QY 486 TTTSTIKLNFDEWSSIASSSRGISGHSCKSGQE---ESFVQSSVQPPGDSKTAPE 541
 Db 424 NTPPS-PAGPTENRENTANENTTLSPVEPTENRETANENTTSPAEPTENGQRTPT 481
 QY 542 E----SSEDVTKYQGVSAEN-----PVENHINITQSDKFTAKPLD-SNSGERNDL 587
 Db 482 EKTTSAAEPTHEGERTPLANENTTSPAPTEPNER-ERTANEKTTSPAPTEPENGDRTP 540
 QY 588 NLDRCGVPESASEKAKEPETSQTSATNENNNPEPQFQTEATGPPSAHETSTR 647
 Db 541 ANEKT--TPSLAEPTEGQRTPFANEKTTSSAEPTEHEERTPLANENTTSPAPTEP 598
 QY 648 D-SALQDTSDDDDPVLIPGARYRAGPGDRRSAVA 681
 Db 599 ERTANENTTSP-----AGPTENRENTA 621

RESULT 16

US-09-839-996-6
 ; Sequence 6, Application US/09839996
 ; Publication No. US20030009010A1
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Gene III, Joseph W.
 ; Falkow, Stanley
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration
 ; Protein
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/839,996
 ; FILING DATE: 20-Apr-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/296,791


```

; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

Query Match          3.9%; Score 175.5; DB 9; Length 1848;
Best Local Similarity 21.6%; Pred. No. 0.018;
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;

QY 250 NKKSCRV-----TSLCYSDGQILVSYSDIYLFDPKDDTARELKTPTSAEERR 299
Db 920 NKKSNKVVVNSATGFTLQVADTKGEP-----NNHLLTDFDASNATNNLEVTLANGSV 974

QY 300 BELRQPPVKRLRLRGDWS-----DTGPR---ARPESEREDGQSPNVSLMQMSDMLSRW 352
Db 975 D-----RGAKYKLRNVNGRYDLNPEVEKRNQTVDTTNIITPNDI-----1015

QY 353 FEEAQAQNRGRSRPRGTSQSDISTL--PTVPSSPDLEVTAMVDTTPAEQFLQ 410
Db 1016 -QADAPSQSN-----NEELARVETVPPAPATESAIAEQPETPAETAQ 1061

QY 411 P-----STSTMSAQAHSTSPSTESHTPLLSPPDSEQSVASGHHHTHQSDNNNE 464
Db 1062 PAMEETANSTETAPKSDTATQTNPNSE---SVPSETTEKVAENPPQENETVAKNQEQ 1118

QY 465 KLSKPKGTGE-----PVLSLHYSTEGTTTITKLNFTDEWSSASSRGGISGSHCKSGEQ 518
Db 1119 ATEPTQNGEVAKEQDPTVEANTQNEATQSEBK---TEETQTATKSEPTESVTVSENG 1175

QY 519 EESVPOSSVQPPGDSSTKAPESESSESVTKYQGVSAENPVENHINITQSDKFTAK-PL 577
Db 1176 PEKTVSQSTEDKVVVEKEKAKVE--TEETQAPQVTSKEPP-----KQAPPAPEVPT 1227

QY 578 DSNGERNDLNRSCGVPESASSEKAKEPETSDQTSATESATNNNTNPPQFQTEATG 637
Db 1228 DTNAEEAQAALQQTPTTAAAEITSPNSKPAETQOPS-----EKTNAEP-----VT 1274

QY 638 PSAHEETSTRDSALQDQDSDDDPVL-IPGARYRAGPCDRBSAVARIOEFFRRKEREKEM 696
Db 1275 PVVSENTATQTEETAKVEKETQEVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 1333

QY 697 EELDTLNIIRP 707
Db 1334 NVLTKNVGE 1344

RESULT 17
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes

```

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match          3.8%; Score 174; DB 10; Length 2368;
Best Local Similarity 21.6%; Pred. No. 0.03;
Matches 103; Conservative 61; Mismatches 181; Indels 132; Gaps 19;

QY 281 DKPDQDARELKTPTSAEERELRQPPVKRL--RLRGDWSDTGPRARPESEERERDGSQSPN 338
Db 1961 DNNDATANS-----NANATPENTGQPNVTSTONANATSTSTTTNNQNDAAATGETTATSAN 2016

QY 339 VSLMQMSDMLSRWEAQAQNRGR--SRPRGTSQSDISTLPTVPSSPDLEVTSET 397
Db 2017 SAT-----DDANDKPOANNSSADTSTNSPTMDNDVTSKPEVESTNNGTTDKP 2065

QY 398 AMEVD--TPAEQFLOPSTSTMSAQAHSTSPSTESHTPLLS-----439
Db 2066 ATEADNATPAES---ATNNSTTTATNENAPTATATPTASTASTAGAESADSKDNASVN 2121

QY 440 -----PDSEQRQSVASGHHHTHQSDNNNEKLSKPKGTGE 474
Db 2122 DSKQNAEVNNSAESOSTNGKVAQPKSENKAKAEKGRDSTNGSMVESTETL-PSADITE 2180

QY 475 PVLSLHYS--TEGTTTSTIKLNFTDEWSSASSRGGISGSHCKSGEQEESFVQ--SSVOP 530
Db 2181 PKVSSNTSKDEESTTSQTDABQHNSDINVASNEAD-----KSEGNVDTDVSNNKPSKPK 2235

QY 531 PEGDSETKAPESESE-----DVTYKQGVSAENPVENHINITQSDKFTAKPL----DSN 580
Db 2236 SEAKDKATSTEDSQADMATADTKDNQASIGATADVNN--KATQDGNANASPATVSKGSN 2293

QY 581 SGERNDLNRSCGVPESASSEKAKEPETSDQTSATESATNNNTNPPQFQTEATGPSA 640
Db 2294 SANQDMLNVTK-----TKENKAKASQAQGVKNKPKQ 2325

QY 641 HEETSTRDSALQDQDSDDDPVLIPGARYRAGPCDRBSAVARIOEFFRRKEREKEME 697
Db 2326 QAKT-----LPDTGMSHND--LPYAEIALGAG-----MAFLIRRTTKDQO 2365

RESULT 18
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.

```


APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 1421
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-13

Query Match 3.7%; Score 169; DB 10; Length 1421;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 136; Conservative 94; Mismatches 241; Indels 240; Gaps 32;

QY 160 CGEDGT-----VRFD-----TRIKTSCKEDCKDDILINCRAATSV- 198
Db 615 KREDIENIQPFRFSEMGDDYCDQTKMIETLVECKEPCDD---NCKRCKNSYKE 671
QY 199 -ICPPIPYLAAGCSDSSVRIYDRMLGTRATGNAGRTTGMVAF-----IPSHLNKS 253
Db 672 WISKKEEY-----NKAQXQYQKG-----NNYK-----MYSEFSIKPEVILKYS 715
QY 254 CRVTSICVSEGOEILVSYSSD-----YLFDPKDDTARELKTPSAEE-- 297
Db 716 EKCSNLNFEDEFKEEL---HSDYKNKTCMCPVKDVPISIRNNQTSQEAYPEESTETA 772
QY 298 REELRQPPVKRLRGDWSDTGPARPESRERDGEOSPVLQMDSML-----SR 351
Db 773 HRTETP-----DTERKNQEPANKDLKPKQPSVGEN---GTDLLQEDLGGR 816
QY 352 WFEASEVAQNRG--RGRSR-----PRGGTSQSDIST-----LPTVP 387
Db 817 SEDEVTOEFGVNHGIPKGEDQTLGSKDAIPNIGETGISTEERHEGHKNQALSTSV 876
QY 388 SSPDLEVSETAMEVDTPAEQLOPSTSTMSAQAHSTSPTESPHST--PLLSSPDSEQ 444
Db 877 DEP--ELSDTLQLHEDTKENDKLPLESTI-----TSPTES-----GSSDPEETPS 920
QY 448 VE--ASGHHTHSDNNNEKLSKPCTGEPVLSLHYSTEGTTSTIKLFTDEWSSIAS 491
Db 921 ISEGPKEQKRDSDLSKLSVSPENSRP-----ETDAKDTNLLKLGVDVIMPKA 974
QY 492 -----KLNFTDEWSSIASSSRGIGSHCKSEGOEESFVQSSVQPPGDSKAPES 543
Db 975 VIGSPNDINVTQGDNIS---GVNSKPLSD-----DVRPKNHEEYKEHTSN 1020
QY 544 SEDV-----TKYQGVSAENPVENHINITQSKFTAKPLDSN 580
Db 1021 SDNVQSGGIYVMNVKELKDTLENPSSSLDGAHELSNPILSSDDQMSNTPOPLDNT 1080
QY 581 SGE-----RNDLNLDRSCG-----VPEBSASSEKAKEPETSQDTSATNENN 624
Db 1081 SEETTERISNNEYKVNREGERTTKYVEDIVLKSHMNRSDGDELXDENSLSTVNDES 1140
QY 625 TNPEPQFQTEATGSAHEETSTRDSALQDQDSD-----DDPVLIPGARVAGPGD 675
Db 1141 EDAEAKMKGNDTSEMSHNSQIHESDQKNDKTVGDLGTHVONEISVP-----1190
QY 676 RRSAVARIQEFRRRKRNK-----EMEELDTLNRPLVYKMYKGHNRSTM 722
Db 1191 ---VTGEIDEKRESKESKIKHAEERLSHTDINH-----INPEDRNSNTL 1233

RESULT 21
US-09-815-242-10932
; Sequence 10932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10932
LENGTH: 1031
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10932

Query Match 3.7%; Score 167.5; DB 10; Length 1031;
Best Local Similarity 23.8%; Pred. No. 0.025;
Matches 67; Conservative 48; Mismatches 122; Indels 45; Gaps 9;

QY 387 PSPDLEVSETAMEVDTPAEQLOPSTSTMSAQAHSTSPTESPHST--PLLSSPDSEQ 444
Db 718 PAAQKRINQLTQTITALLLVKSTETSTNTSESSTSTSTSTSTSTSTSTSTST 777
QY 445 ROSVEASGHHTHHQSDNNNEKLSKPCTGEPVLSLHYSTEGTTSTIKLFTDEWSSIAS 504
Db 778 SSTSESST 827
QY 505 SSRGIGSHCKSEGOEESFVQSSVQPPGDSKAPESSESDVTKYQGVSAENPVENHI 564
Db 828 SS-----TSSTTNBSSSTSESST--PSTTSESSTSTSTSTSTSTSTSTSTST 879
QY 565 NITQSDKFTAKPLDSNGERDNLDRSCGVPEESASSEKAKEPETSQDTSATNENN 624
Db 880 SET-----SNTNESNTPSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 928
QY 625 -----TNPEPQFQTEATGSAHEETSTRDSALQD 653
Db 929 QSKGQSVIYAVESNQDPNDAQSNKPSA-KASQTKESVAEN 969

RESULT 22
US-09-815-242-13057
; Sequence 13057, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard


```
Db 676 KKEYNKQAKYQYQKGNMYSEFKSIKPEVYLKKYSEKSNLNFDEFKEE-----730
QY 374 GTSQSDISTLPTVPSPDLEVSSTAMEVDTPAQFQLPSTSTMSAQAHSTSP--TESP 431
Db 731 --LHSDYKKNKIM--CP--EVKDVPIIIRNNEQTQEAPEENTIAHRTSPISSEGP 784
QY 432 HSTPLSSPDSEQORVEASGHHTHSDNNNEKLSKPKGTGTPVLSLHYSTEGTTSTI 491
Db 785 -----KGNEKQERDDDSLSKISVSPNSRP-----ETDAKDTNLS 819
QY 492 -----KLNFTDEWSSSTASSRGRIGSHCKSEGEQSEFVPOSSVQ 529
Db 820 LKLGVDVLSMKAVIGSPNDNINTEQDNLS-----GVNSKPLSD-----DVR 865
QY 530 PPEGSETKAPESSESDVTKYQGVSAENPVENHINITQSDAKFTAKPLDSNSG-----ER 584
Db 866 PKKLEEDQNSDESETVNH--ISKSPINNGDDSGSATVSESSSNTGLSIDDDR 922
QY 585 NDLNDRSCGVPESASSEKAKPETSDQSTES-----ATNENNTNPEQFQTEATG 637
Db 923 NGDTFVRT-----QDANTEDVIRKENADKDEKGADEERHSTSESLSPEEKMLTDNEG 978
QY 638 PSA--HEE-----TSTRDSA-----LQDT-----DD-----656
Db 979 GNSLNEEVEKHTSNDVQSGGIVNMVVEKELKDTLENPSSLDDEGKAHEELSEPMLS 1038
QY 657 SDDPVLIPGARYAGPDRRSARAVARIQEFFRR--RKERKEMEELDTLIRPLVKMYK 714
Db 1039 SQDMSNTP-----GPLDNTS-----EETTERISNNEKYVNEREDERTLTREYEDIVLK 1087
QY 715 GHNSR-----TWIKANFWGANFVMSGDCGHIFWDRITABHMLLEAD 760
Db 1088 SHMNRSDGELYDNSDLSTVNDSE--DAEAKMKGNDTSEM--SHNSQH-----IESD 1139
QY 761 -----NHVNCLOPHFPDPTLASSGDYDI-----KWSPLEESRIFN--798
Db 1140 QKNDKMTVGLDGLTHVQNEIS-----VPVTGEIDELKRESKSI--HKAEEERLSHTD 1192
QY 799 --RKLADDEVITRNELMLETRN 818
Db 1193 IHKINPEDRNSNTLHKDIRN 1213
```

RESULT 24

```
US-10-025-380-1065
; Sequence 1065, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedick Thomas S.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1065
```

; LENGTH: 957

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-025-380-1065

Query Match

Best Local Similarity 3.6%; Score 164.5; DB 9; Length 957;

Matches 132; Conservative 81; Mismatches 301; Indels 149; Gaps 28;

QY 126 EQDAENR--QCQTCYCHYTYEIMVNP-----DPYFLSC--GEDGTVMFDTRIKT- 175

Db 204 ESDTSGREESTSHSSTHTTSSAPSTTSALVEEPTSYHSPGSTATHTFPDSSTSG 263

QY 176 -----SCTKECKDDILLNCRRAATVAICPPIPYLLAVGCDSSVRIYDRMLGTRA 228

Db 264 RSESTASHNQDATGIVLPA--RSTSVLLGESTVSPISGSMETTA-----LPGSTT 316

QY 229 TGNVAGRTGMVARIPSHLNKSCRVT-----SLCYSEDGQELIYSYS---SDIYILFD 281

Db 317 TPLGSEKST-----FHSSPRPATLSPASTTSSGVSEESTTSHSRPGSTHTTAF- 367

QY 282 PKDDTARELKTPSAERREELRQPP--VKRLRLRGWSDTGPRAPESEERERGEQSPV 339

Db 368 --PDSTT-----TPGLSHSTSHSPGSTDTLLPASTTTSGP--SQESTTSHSPGSTDT 420

QY 340 SLMQMSDMLSRWFEEASEVAQSNRG-----RGRSRPRGGTSQSDIST 382

Db 421 ALSPGSTTALS--FGOESTTFHSSPGSTHTLFPDSTTSSGIVEASTRVHSSTGSPRTLL 478

QY 383 LPTVPSPDLEVSETAMEVDTPAEQFLOPSTSTMSAQAHSTSPTESPSTPL-----437

Db 479 SPASSTSPGLOGESTAFQTH--PASTHTTPTSTATAPVEESTTYHRSPTSPTTHFPAS 537

QY 438 -----SSPD-----SEQORVEASGHHTHSDNNNEKLSKPKG--TCEPV 476

Db 538 STTSGHSEKSTTFHSPGDASGTTSPSAHSTTSGRGESTTTSRISPGSTTEITLFGSTTTPG 597

QY 477 LS-----LHYSTEGTTSTIKLNFDEWSSSTASSRGRIGSHCKSE---GOEESFVPOSSV 528

Db 598 LSEASTTFYSSPRSPPTTL-----SPASMTSLGVEESTTSRQPGSTHSTVSPAST 649

QY 529 QPP-----EGDSEYKA--PEESEDVTKYQGVSAENPVENHNI-----566

Db 650 TTPGLSEESTVYSSPGSTETTVFPRSTTSYVRGEPTTFHSPASTHTLTEDSTTS 709

QY 567 ---TQSDKFTAKPLDSNSGERNDNL-----DRSCGVPEESASSEKAKEPESDQSTESA 619

Db 710 GLTEESTAPGSPASTQTGLPATLTADLGEESTTTPSSSGSTGTTLSPARSTSLGVE 769

QY 620 TNENNTNPEQFQTEATG-----PSAHEETSTRDSALQDSTDSDDPVLIPGARYRAGPD 675

Db 770 STPSRLSPSSTETTLPGSPSTTTPSLSEKSTT-----FYTSRSPDATLSPATTTSSGVSE 824

QY 676 RES 678

Db 825 ESS 827

RESULT 25

```
US-09-922-217-1065
; Sequence 1065, Application US/09922217
; Patent No. US2002007641A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yudi
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
```


Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5816
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 3.6%; Score 164.5; DB 10; Length 2478;
Best Local Similarity 21.1%; Pred. No. 0.12;
Matches 97; Conservative 71; Mismatches 187; Indels 105; Gaps 21;
QY 281 DPKDDTARELKTPSAERREELRPPVKRLRLRGDSMTGPRAPRESERERDGEQSPNVS 340
DB 2078 DNNDATANS-----NANATPENTGQPNVSETTANGK-ADASP-TTPNNSDAATGETTATSA 2131
QY 341 LMQRMSDMLSRWFEEASEVAQSNRGRG-RSRPRGGTSQSDISTLPTVPSPDLEVSSTAM 399
DB 2132 T-----DDANDKQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPVT 2179
QY 400 EVD--TPAQFLOPSTSTMSAQAHSTSPESP--HSTPLLSPPDSEQSQVEASGHHT 455
DB 2180 ETDNATPAESTTNNSTTTATNENAPTGTATPTASTAEASADSADKSNASVNDKQNA 2239
QY 456 H---HSDNNNEKLSKPKCTGEPVLSLHVSTGTTTSTIKLNFTEWSSIASRRGIGS 511
DB 2240 EYVNSAESOSTNDKVA-QPKSENKAKAEKDGSDSTNQSMVE-STTETLPADIETPNVPS 2297
QY 512 HCKSEQEESFVPOSSV-----QPPEGDSE-----TKAPEESSEDVTK 549
DB 2298 N-TSKDKESTTNTQDAGOLKSETNVASNEADKSPSKADTEVSNKPKTSASSEAKKMTS 2356
QY 550 YQEGVSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPPEESASSEKAKEPE 609
DB 2357 -----TNVSKD-----DTATADTND--TQKSVG-----SAANKATQ-- 2387
QY 610 TSDQSTESATNENNTPEQFQTEATGSAHE-----ETSTRDSALQDRTDS 657
DB 2388 -NDGANASPATVSGNSANSQDMLNVTNTDDHOAKTKSAQQGKYNKAKQOAKTLPTDGM 2446
QY 658 DDDPVLPGARYAGCDRRSAVARIQEFFFFRRKERKEME 697
DB 2447 HNDD--LPTAEALGAG-----MAFLIRRTTKDQ 2475

RESULT 28
US-09-815-242-12967
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12967
LENGTH: 2478
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12967

Query Match 3.6%; Score 164.5; DB 10; Length 2478;
Best Local Similarity 21.1%; Pred. No. 0.12;
Matches 97; Conservative 71; Mismatches 187; Indels 105; Gaps 21;
QY 281 DPKDDTARELKTPSAERREELRPPVKRLRLRGDSMTGPRAPRESERERDGEQSPNVS 340
DB 2078 DNNDATANS-----NANATPENTGQPNVSETTANGK-ADASP-TTPNNSDAATGETTATSA 2131
QY 341 LMQRMSDMLSRWFEEASEVAQSNRGRG-RSRPRGGTSQSDISTLPTVPSPDLEVSSTAM 399
DB 2132 T-----DDANDKQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPVT 2179
QY 400 EVD--TPAQFLOPSTSTMSAQAHSTSPESP--HSTPLLSPPDSEQSQVEASGHHT 455
DB 2180 ETDNATPAESTTNNSTTTATNENAPTGTATPTASTAEASADSADKSNASVNDKQNA 2239
QY 456 H---HSDNNNEKLSKPKCTGEPVLSLHVSTGTTTSTIKLNFTEWSSIASRRGIGS 511
DB 2240 EYVNSAESOSTNDKVA-QPKSENKAKAEKDGSDSTNQSMVE-STTETLPADIETPNVPS 2297
QY 512 HCKSEQEESFVPOSSV-----QPPEGDSE-----TKAPEESSEDVTK 549
DB 2298 N-TSKDKESTTNTQDAGOLKSETNVASNEADKSPSKADTEVSNKPKTSASSEAKKMTS 2356
QY 550 YQEGVSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPPEESASSEKAKEPE 609
DB 2357 -----TNVSKD-----DTATADTND--TQKSVG-----SAANKATQ-- 2387
QY 610 TSDQSTESATNENNTPEQFQTEATGSAHE-----ETSTRDSALQDRTDS 657
DB 2388 -NDGANASPATVSGNSANSQDMLNVTNTDDHOAKTKSAQQGKYNKAKQOAKTLPTDGM 2446

QY 658 DDDPVLIFGARYRAGGRRSAVARIQEFFFFRRKKERKEME 697
Db 2447 HND--LPVAELALGAG-----MAFLIRRTKKDQ 2475

RESULT 29

US-09-765-272-94
; Sequence 94, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: P8340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-09-765-272-94

Query Match 3.6%; Score 164; DB 10; Length 278;
Best Local Similarity 22.2%; Pred. No. 0.0074;
Matches 66; Conservative 50; Mismatches 117; Indels 64; Gaps 10;

QY 387 PSSPDLEYSETA-----MEVDTPAEQFLOPSTSSMSQAHSSTSPTE 429
Db 9 PESDPTVSDRGEPOQVAPLPEYKGNIEQVPEPVEKTEKQGEKTEEV-----PVK 61
QY 430 SPHSTPLSSPDSEORQVSEAGHHTH--HOSDNNKLSKPKGTGEPVLSLHSTEGTT 487
Db 62 PTEETPVNPNEGTTBGTGIEQAEVNPQAEESTTNSKRVSP-----DTSSKN 108
QY 488 TSTIKLNTDWMSSSTASSRGISGCKSEGOEESVPQSSVQPPGDSSETKAPESSESDV 547
Db 109 TGEVSSNPSSDTSVGSNGKPEHNDKKNSEKT--VEEPVNPNEGTVGEGTSNQETKPV 167
QY 548 TKYQGVSAENPVENHINITQDKFTAKPLDSNSGSDKDLNDRSCGVPESASSEK--- 604
Db 168 QPAEE-----TQTN--SGKIANENTGEVSNKPSDSKPPV-EESNQPEKNGT 210
QY 605 AKPEPSTQSTESATNNNNPNPEQFQTEATGSPAHEETSTRSALQDSTDSDDDP 661
Db 211 ATKPENSNGTTSN-----GQTEPEPSNGNSTEDVSTNSNTSNGNEEIKQENELDP 263

RESULT 30

US-10-077-111-10
; Sequence 10, Application US/10077111
; Publication No. US20020187492A1
; GENERAL INFORMATION:
; APPLICANT: Todderud, C. Gordon
; APPLICANT: Finger, Joshua N.
; APPLICANT: Rillema, Jill
; TITLE OF INVENTION: TBA
; FILE REFERENCE: 3053-4114US2
; CURRENT APPLICATION NUMBER: US/10/077,111
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/294,181
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/269,366
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1356
; TYPE: PRT
; ORGANISM: Podospora anserina
; FEATURE:
; OTHER INFORMATION: beta transducin-like protein encoded by the
; OTHER INFORMATION: het-e-1 gene
US-10-077-111-10

Query Match 3.5%; Score 160.5; DB 9; Length 1356;
Best Local Similarity 20.9%; Pred. No. 0.096;
Matches 97; Conservative 65; Mismatches 182; Indels 121; Gaps 21;

QY 46 TLNVHDCVNTICWNTDTEYILSGSDTKLVISNPSYRKVLFTIRSHRANIFSAKFLPC 105
Db 920 TLEHGGRVQSVAFSPDQGVASGSDHTIKIWDAAAGTCTQTL-EGHGSSVLSVAFSP- 977
QY 106 TNDKQIVSCSGDGVIFYTNVEQDAETNQCQFTC-----HYGTIYEIMTVNDPYTFLS 159
Db 978 -DQQRVASGSGDKTIKIWDTAGS-----TCTQTLEHGGSVMSVAFSP-DGQRVAS 1026
QY 160 CGEDGTVRWFDTRIKTSCTKEDCKDDILINCRRAATVAICPPIPYILAVGSDSSVRIY 219
Db 1027 GSDDKTIKIWDTAGSCT-CTQ-----TLEHGGSVQSVVFPDQQRVASGSDHTIKIW 1078
QY 220 DRMLGTRATGNAGRTTGMVARFIPSHLNKSCRVTSLCYSEDGQELVLSYSDIYL 279
Db 1079 D-AVSGT-CTQTLEHGSDS-----VMSVAFSPDQQRVASGSDGTIKI 1119
QY 280 FDPKDDTARELKTPSAEERRELRQPPVKRLRLRGDWSDTGPRARPESEREDGEQSPNV 339
Db 1120 WDAASGTC-----TQTLEHGGSVMSVAFSPDQQRVASGSDGTI 1159
QY 340 SLMQMSDMLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLTPTVPSSPDLVSETAM 399
Db 1160 KI-----WDAASGTCQTQLEG-----HGWVQS-----VAFSPD----- 1188
QY 400 EVDTPAEQFLOPSTSSMSA--QAHSTSPSTPESHSTPLLS---SPDSEQROSVASGHH 454
Db 1189 -----QQRVASGSSDKTIKIWDTAGSCTQTL-EGHGGSVQSVAFSPDQQR-----VASG-- 1237
QY 455 THHQSDNNKLSKPKGTGEPVLSLHSTEGTTTSTIKLNTFDEW 499
Db 1238 ---SSDNTIKIWDTAGSCTQTLNV-----GSTATCLSFDTNAY 1274

RESULT 31

US-09-815-242-5904
; Sequence 5904, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Kari L.
; APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-03-21
PRIORITY FILING DATE: 2000-05-23
PRIORITY FILING DATE: 2000-07-27
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-04-24, 578
PRIORITY FILING DATE: 2000-10-23
PRIORITY FILING DATE: 2000-11-27
PRIORITY FILING DATE: 2000-12-22
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5904
LENGTH: 502
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5904

Query Match 3.5%; Score 160; DB 10; Length 502;
Best Local Similarity 21.9%; Pred. No. 0.028;
Matches 85; Conservative 56; Mismatches 148; Indels 100; Gaps 12;

QY 312 LRGWSDTGPAPRPERERDEQSPNVLQRMDSMLRWFEAEVAQSNRGRSRP 371
Db 53 LSGTSESESDTSSESSEKSD-STSMISMSQSTS----- 86
QY 372 RGGTSQSDISTL-PTVPSSPDLEVTAMEVDT-PAEQFLQPTSTSTMS----- 418
Db 87 -GSTSTSTSLSDSTSTSLASMSNGVDNSASASASSTSTSTSESQSTSSYT 145
QY 419 ----AQHSTSPSPSTPLLSGPDSEQSQSVASGHHTHQSDNNNEKLSKPGTGE 474
Db 146 SQSTSQSESTSTSLSDSTSIKSTSQSVSTASLSGSESDSQSISTASBESTSE 205
QY 475 PVL-----SLHYSTEGTTSTIKLNTDEWSSIASSSRGIGSHCKSEGOEESFVQSSVQ 529
Db 206 SASTSLSDSTSTSGSASTSTSLN-----NSASASEDLSTSLSDS-----TSASMQ 254
QY 530 PPEGDSSTKAPES-----SEDTKYQEGVSAENPVENHINITQSDKF 572
Db 255 SSESQSTASLSDSTSTSNRMSTTASLSTSVSTSEGSTSE-----STSESST 307
QY 573 TAKPLDNGSRNDLNLDRSGVPEESASEKAPETSDQSTESATN-----ENNINPE 628
Db 308 STSLSDSQSTERS-----TSASGASTSTSTSDSRSTASSTSMRTSTSDSQ 355
QY 629 PQFQTEATGPAHETSTRDALQTDSDS 657
Db 356 SMSLSTSTSTSMDSSTSLSDSVSDSTSDS 384

RESULT 32

US-09-801-368-408
Sequence 408, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Cali, Brian
APPLICANT: Hecht, Peter

APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801.368
CURRENT FILING DATE: 2001-03-07
PRIORITY FILING DATE: 2000-01-19
PRIORITY FILING DATE: 2000-01-19
PRIORITY FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 408
LENGTH: 713
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-408

Query Match 3.5%; Score 156.5; DB 10; Length 713;
Best Local Similarity 19.6%; Pred. No. 0.072;
Matches 111; Conservative 92; Mismatches 225; Indels 137; Gaps 23;

QY 286 TARELKTPSAEERELRQPPVKRLRLRGDWSDTG-PRARPESERERDEQSPNVLQMR 344
Db 65 THRKMKDAYEEE-----IKHLKLEQRDQIASLTVOQQRQOQQOQQVQHLQQ 115
QY 345 MSDMLSNWFEASVAQ-----SNRGRG-----RSRPRGGTSQSDISTL 383
Db 116 QQQLAAA--SASVPAQPPATTSATATPAATTTGSPSAFPVQASRPNVLSQPTTL 173
QY 384 PTVPSSPDLEVTAMEVDTPAEQFLQPTSTSTMSAQAHSTSSPTSEPHST-PLLSPPDS 442
Db 174 PVYSSNAQQQLPQOQLQOQQOQQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPES 233
QY 443 EQQSVASGHHHTHQSDNNNEKLSKPGTGEVLSLHYSTEGTTST-IKLNTDEWSS 501
Db 234 TLKET-----EPENNTSKIN--DTGSA-----TTATTTTATETETKPKEDAT 275
QY 502 IASSRIGSHCKSEGOEESFVP-----QSSVQPP-----EGDSEI--KAPESESDYTK 549
Db 276 PASLH-----QDHYLVYNQRANHSKPIPPFLDLSDSQSVDPALKQTNDYIL 324
QY 550 YQEGVSAENPVE-----NHINITQSDKFTAKPLDNGSRNDLNLDRSGVPEESASSEK 604
Db 325 YNPALPREIDVELHKLSDHISVCCVKF-----SNDEYLATGCNKTKTYVRVSDGSLV 378
QY 605 AKETPESQSTESATNNTNPEPQFQTEATGPAHETSTRDALQ-----DTDD 656
Db 379 ARLSDDSAANHRNSITENTTTTNDNTMTTTTTTTTTTTTAMTSAELAKDVENLTSS 438
QY 657 SDDDPVLI-----PGARYA-GPDRRSARVATQEFFRRKEREKEMEELDTLNRPLV 709
Db 439 SPSSDLYIRSVCFSPDGKFLATGAEDR---LIRLWDIENRK-----I 477
QY 710 KMVYKGRNSRTMIKEANFWGANF-----VMSGSDCHGFIWDRHTAEHLMLLEADNHV 763
Db 478 VMILQGH-----EQDIYSLDYFPDGKLVSGSGDTRVRLWDLTQCQSLTSLDEGV 529
QY 764 VNCLOPHFPDPIASSGIDYDIKIW 788
Db 530 TTVAVSPDGKYGIAAGSLDRAVRVM 554

RESULT 33
US-09-820-843A-12

; Sequence 12, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 505
; TYPE: PRT
; ORGANISM: H. pylori
; NAME/KEY: misc feature
; OTHER INFORMATION: poly E-rich protein
; NAME/KEY: misc feature
; OTHER INFORMATION: g1|2313421
US-09-820-843A-12

Query Match 3.4%; Score 155.5; DB 9; Length 505;
Best Local Similarity 18.2%; Pred. No. 0.053;
Matches 74; Conservative 73; Mismatches 152; Indels 107; Gaps 15;
QY 239 GVARFIPSHLNKSCRYTSLCYSDGQILVSYSSDYIYLPDKDDTARELKTSPSABER 298
Db 145 GDLEALVQEPNNEBQLLPTLNDQKEKEVKE-----EKEEVEKEEKEEVEKEE 194
QY 299 REELRQPPVKRLRGDWSDTGPRARP----- 325
Db 195 KEVKETQEEKKPKDDTQGETLKDKEVSKELPAQELPKEETQPDPIKEETOEN 254
QY 326 ESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNRGRSRPRGCTSQSDISTLPT 385
Db 255 KEKQEKTDSPSAQELAMQELV-----KEIQENSGQ-----ENKKTQES 297
QY 386 VPSPDLVSET-----AMEVDTPAEQFLOPSTSTMSAQAHSTSP--TESPHSTPL 437
Db 298 AEIPQDKIEQVVTETKTAQAELEVPKEK--TQSAEALQETAHELEKOEIAETPDQVEIP 356
QY 438 SSPDSEQSQVSEASGHHTHQSDNNNEKLSKPGTGPVLSHYSTEGTTTSTIKLNTD 497
Db 357 QSQDKEV-QELEIPKEETOENTETPDVETPQE-----KETQED 394
QY 498 EWSSTASSRIGSGHCKSEGESEFVPOSSVQPPGDSSETKAPESSESDVTKYQEGVSAE 557
Db 395 HYESIEDIPEV--MAKANGEELPFLNEAVAKIPNNENDTETPKESVTETSKNEN--NTE 450
QY 558 NPVENHINITQSDKFTAKPLDSNGERNDLNL--DRSCGVPEESASS 602
Db 451 TPQKE-----ESDK--TSSPLELR-----LNLQDLLKSLNQESLKS 485

RESULT 34

US-09-815-242-11317
; Sequence 11317, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11317
; LENGTH: 505
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11317

Query Match 3.4%; Score 155.5; DB 10; Length 505;
Best Local Similarity 18.2%; Pred. No. 0.053;
Matches 74; Conservative 73; Mismatches 152; Indels 107; Gaps 15;
QY 239 GVARFIPSHLNKSCRYTSLCYSDGQILVSYSSDYIYLPDKDDTARELKTSPSABER 298
Db 145 GDLEALVQEPNNEBQLLPTLNDQKEKEVKE-----EKEEVEKEEKEEVEKEE 194
QY 299 REELRQPPVKRLRGDWSDTGPRARP----- 325
Db 195 KEVKETQEEKKPKDDTQGETLKDKEVSKELPAQELPKEETQPDPIKEETOEN 254
QY 326 ESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNRGRSRPRGCTSQSDISTLPT 385
Db 255 KEKQEKTDSPSAQELAMQELV-----KEIQENSGQ-----ENKKTQES 297
QY 386 VPSPDLVSET-----AMEVDTPAEQFLOPSTSTMSAQAHSTSP--TESPHSTPL 437
Db 298 AEIPQDKIEQVVTETKTAQAELEVPKEK--TQSAEALQETAHELEKOEIAETPDQVEIP 356
QY 438 SSPDSEQSQVSEASGHHTHQSDNNNEKLSKPGTGPVLSHYSTEGTTTSTIKLNTD 497
Db 357 QSQDKEV-QELEIPKEETOENTETPDVETPQE-----KETQED 394
QY 498 EWSSTASSRIGSGHCKSEGESEFVPOSSVQPPGDSSETKAPESSESDVTKYQEGVSAE 557
Db 395 HYESIEDIPEV--MAKANGEELPFLNEAVAKIPNNENDTETPKESVTETSKNEN--NTE 450
QY 558 NPVENHINITQSDKFTAKPLDSNGERNDLNL--DRSCGVPEESASS 602
Db 451 TPQKE-----ESDK--TSSPLELR-----LNLQDLLKSLNQESLKS 485

RESULT 35

US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

Db 481 RRG-----SISVSXPKP-----LSSFKMSLRNWWTRTPSSPPV 514
QY 510 ---GSHCKSEGESESVPOSSVOPPGDSETKAPESSESDVTKYQGVSAENPVENHIN- 565
Db 515 TTPASETKISSPRKALIPVS-----OKSQADACSESRRVRRLDSCLLESVKQKCVKS 569
QY 566 ---ITQSDKFTAKPLDLSNGERNDLNRSCGVPESASSEKAKEPETSDOISTESATNE 622
Db 570 CNCVTELD-----GQAESRLDLCCLSGTOGEVLSDQSEGPTKSKTEGAGTSIS 618
QY 623 NNTNPEPQTEATGP-----SAHEETSTRD-----SALQDQDSD--D 659
Db 619 EPPSPVPSYASECGPLPLRPGCGSEWCKENSSPENKNWLLAIAAKRAENSPPRS 678
QY 660 DVLIPGARYRAG---PGD---RRSAVARIQEFFRRRKE 692
Db 679 PSSQTPSSRRSQSGKTPGPVTTIPPSMRKICTYFRKKTQ 717

RESULT 37

US-10-108-605-195
; Sequence 195, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Klm
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 195
; LENGTH: 980
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-195

Query Match 3.4%; Score 152.5; DB 9; Length 980;
Best Local Similarity 20.5%; Pred. No. 0.19;
Matches 92; Conservative 70; Mismatches 193; Indels 93; Gaps 17;

QY 256 VTSLC--YSEDDGOEIL--VSYSSDXIYLPDPKDDTARELKTPTSAEERREELRQPPVKRLR 311
Db 599 IITLCVYFVENFSKIFDEIYVQTFYKLNRYDQYQDLKDRDKMENRTDGGLPITR--- 655
QY 312 LRGDWSDTGPRAPESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQ-----SNRGR 366
Db 656 -----SGRFR-----RDQRQ-----MEEEREM---WNEEDDFTFEEIDTYNNVMK 693
QY 367 GRSRPGGTQSQSDISTLPVPSPPDLEVSETAMEVDTPAEQFLQPTSTMTMSAQAHSTSS 426
Db 694 SVSEKNGPQTQNKQSSPPHSTP-----HSGILGSLSTTASSTATSATSS 738
QY 427 --PTSPHPTPLLSPPDSBORQSV--ASGHHHTHOSDNNNEKLSRKPCTGEP-VLSLHY- 481
Db 739 GAPVAGSGSSPPAISADEQTQAAVHLAAALQHQQOQQOQQOQNPPQQTQPEIAELOQO 798
QY 482 --STEGTTTSTIKLNF-----TDWESSIASSSRGIGSHCKSEGESESVPOSSVQPP 531
Db 799 LSSVEAPOQSOELESQAASASPTSSSSSLSEASTSSSSASSSSSSSSSSSPPGSSAAS 858
QY 532 EGDSETKAPEESEDVTKYQGVSAENPVENINTQSDKFTAKPLDLSNGERNDLNR 591
Db 859 LCDSATVAAVAASQFLSTIATAMAA-----SVTAAATNSSP-----SISP 899

QY 592 SCQVPESASSEKAKEPETSDQTSATNNTNPEPQTEATGSPSAHEETSTRDSAL 651
Db 900 APAVSPDIENADAQLPPSDDASSPASGEQDANS-----TEGTSEADKTTAKRGLVD 952
QY 652 QDTRDDDDPVLIPGARYRAGPGDRSA 679
Db 953 YESDSGEDD---YEEDEYSEGPOAQKRA 977

RESULT 38

US-09-738-877-3
; Sequence 3, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynnne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-877-3

Query Match 3.4%; Score 152.5; DB 10; Length 1781;
Best Local Similarity 20.7%; Pred. No. 0.42;
Matches 141; Conservative 86; Mismatches 249; Indels 205; Gaps 33;

QY 281 DPKDDTARELKTPTSAEERREELRQPPVKRLRGLDWSGTGPRAPESERERDGEQS 336
Db 339 DGKAEVASEKLTASEQAHQPEAPESAHEP-----RLSAEYE-----KVLPSEQVSGSQG 389
QY 337 PNVSLMQMSDMLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLPVPSPPDLEVSE 396
Db 390 PS---BEKPAPLATEVFEDEKIEVHQEE-----VVAEVHVSVTEERTEQKTEVEE 436
QY 397 TAMEVDTPAEQFL-----QPTSTSS-----TMSAQAHSTSPSPHSTP---LLSPDSE 443
Db 437 TAGSV--PAEELVGMDAEPQAEPAKELYKLTCTCVSGEDPTQGDLSDEKVLSPPEG 494
QY 444 QROSVASGHHHTHOSDNNNEKLSRKPCTGEPVLSLHYSTEGTTTSTIKLNFDEWSSIA 503
Db 495 VYSEVEML-----SSQERMKVQ---GSPLKLLF-----TSTGLKLLSGKKQ----- 532
QY 504 SSSRGIGSHCKSEGESESVPOSSVQPPPE---GDSFKAPESE----- 545
Db 533 KGRKGGD---EESGHTQVPADSPDSQEQKGESSASSPPEEITCTLEKGLAEVQDQ 589
QY 546 -----DVTYQGEVSAENPVENINTQSDKFTAKPLDLSNGERNDLNRSCGVP 597
Db 590 EAEGATSDGEKKREGVT---PWASFKNVTPKKRVRP---SESDELDKVKSATLSS 644
QY 598 -BSASE-----KAKEPETSDOTST-----ESATNNTNPEPQFQ 632
Db 645 TESTASEQEMKGSVEEPKPEPKKVTSVSWEALICVSGSKKARRRSSD----- 698
QY 633 TEATGSPA-----HEESTRDSALQDQDSD-----DPVLIPGARYRAGPGDRSA 679
Db 699 -EGGPKMGGDHQA---DEAGKDKETGDTGLAGSQEHDPQGGSSSPQAGSPTEGEG 754

QY 680 VARIOEPR---RRKERKEMEELDTLNI-----RRP-----LYKMYVKGHR 717
Db 755 VSTWSEFKRLVTPRKSKSKLEEKSEDSIAGSGVEHSTPDTPEFGKESWYSIKKFIPIGR 814
QY 718 NSR-----TWIKANFWANFVMSGDCGHIFIDRHTAHLMLLEADNHVNVNCLQPH 770
Db 815 KRPDQKQEQAPVEDAGPTGAN--EDSDVPAVPLSEYDAVEREKKEAQAQKGAQPE 872
QY 771 P-----FPIPLASSGIDYDIKIW-----SPLEESKIFNRKLA 802
Db 873 OKAATEVSKELSESQVHMAAAVADGTRAATTIERSPSWISASVTEPLEQVEAEAAALT 932
QY 803 DEVITRNELMLEETRTITVP 823
Db 933 EVLER-EVIAEEPPTVTEP 952
RESULT 39
US-09-820-843A-15
; Sequence 15, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820, 843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: M. genitallium
; NAME/KEY: misc.feature
; OTHER INFORMATION: cytoadherence-accessory protein
; NAME/KEY: misc.feature
; OTHER INFORMATION: g1|1046012
US-09-820-843A-15

Query Match 3.3%; Score 151.5; DB 9; Length 1139;
Best Local Similarity 18.3%; Pred. No. 0.27;
Matches 140; Conservative 115; Mismatches 288; Indels 220; Gaps 33;
QY 261 YSEDGQELIVSY-----SSDIYILFDPKDDTARELKTPTSAERR----- 299
Db 106 YDENGWWSGYFENDQWISTKESQPTDENYGF--SDLPPEVKQPVESVEDNYGFDNDLP 163
QY 300 EELRQPPVKRLRLRGDWSGTGPRAR--PESERER-----DGEQSPNVS 340
Db 164 PEVKQPVESVEDNYGF--NDLPPEVKQPVESVWDQPSDDYFAKQPTDENYGFNDLPPEVK 222
QY 341 LMQRMSDLR--WFEASEVAQNRGRGRSRPRGTSQSDI-----SILPTV 386
Db 223 QPESVVDQPSDDHFAKQPESTTDSYFSDLPQPTLDQPSLDHRYQNFDRHLEKPA 282
QY 387 PSSPDLEYSETAMEYDTPAEQFLQPTSTMTGAQAHSSTPSPTPLLSPPDSEOR- 445
Db 283 EQNNYQVGFQDQVQANLDNNEELQPTAEKKVYTDFFESKQAQVDSYQLPI---DIDQD 338
QY 446 QSVEASGHHHT---HQSDNNNEKLSR--KPG--TGEPLVLSLHVS----- 482
Db 339 QTFSSSFETQPTVEQFDQVNSEVNDQFKEITKEPVLESFNNKQDVVETSDLNSENLY 398
QY 483 TEGTTTSTIKLNFTEWSSIASRSGIGSHCKSEGBE-----SFVQSS-----V 528
Db 399 SNNNDATNDSINSEFIQNSINSETASDDVHYESKSEPTHDYKFGSDLSQSNNLSLES 458
QY 529 QPEGDSETKAP-----EESSEDVTKYQEGVSAENPVENHINITQSDKFTAK--PLDSN 583
Db 459 EPVKFNSET--APDAHFESQSEPVQVQYDIYQNEELKPLTLDQSSDDYFAKQPTDENYGF 517

QY 584 RNDNLND---RSCGVPEBSASSEKAKEPE-----TSDQTSTE----- 617
Db 518 DNDLPPEVKQPVESVWDQPSDDHFAKQPESTTDSYFSDLPQPTLDQPSLDHRYQNF 577
QY 618 -----SATNENNTNPEPQFQTEATGTPSAHEETSTDSALQ--D 653
Db 578 HHEELKPAVEBQNNYQVCFDQVQANLDNNEELQPTAEKEVYTDFFESKQAQVDSYQLPID 637
QY 654 TDDSDDDPVLIPGARYRAGPDRKSAVARIQEFFRRRKERKEMEELDTLNIR-----RPL 708
Db 638 TDQDQD-----TTFSSSFETQPTVEQFDQVNSEVNDQFKPE 673
QY 709 VKMYVKGHRNSRTMIKEANFWANFVMSGDCGHIFIDRHTAHLMLLEADNHV--VNCL 767
Db 674 I-----TKPEVLESFNNKQDVVETSNYNNLQKFD-----IQSDNKIITTK 715
QY 768 QPHFPDPI-----LASSGIDYDIKIWSPLESRIENRKLADDEVITRNELMLEETRTITV 822
Db 716 KSPQIPTLPIISFVSNRIEY-----KPVETLALNKESQEQITINS--ITEDSKTLAK 768
QY 823 PASFMLRLASLNH---IRADRLEGDRSEGSGQEN--ENEDEE 860
Db 769 TLSVOLQOQINSLNQSIQVTSVSRVLDKDDQLTINTVNSEDDQ 811
RESULT 40
US-09-801-368-110
; Sequence 110, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busdy, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801, 368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 110
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-110
Query Match 3.3%; Score 151; DB 10; Length 1075;
Best Local Similarity 19.0%; Pred. No. 0.27;
Matches 119; Conservative 80; Mismatches 277; Indels 150; Gaps 21;
QY 70 SDDTKLVLSNPSRRKVLTTIRSGHRANITFSAKFLPCTNDKQIVSCSGDVIFVYTNVEQDA 129
Db 350 TDEVIVIRTPTSEGLITTTTE-----PWT-----GFTTSTSTEMT 386
QY 130 ETNRQOQPT-----CHYGTTYEIMTVPNPDYTFILSCGEDGTVRWFDTRIKTSCTKEDCK 183
Db 387 VTGNGQPTDEVIVIRTPTSEGLITTTTEPWT-----GFTTSTSTETITGNGQP 439
QY 184 DDILINCRRAATSVACIPPIPYLAVGCSDDSVRIYDRMLRMLTRATGNVAGRTGTGMVAR 243

Db 440 TDETIVIRTPTEGLI-----TTTTTPTGTF----- 468
QY 244 FIPSHLNKSCRWTSYCYSEDGQOELVSYSDYIYLFDPKDDTARELKTSPSAFEREELR 303
Db 469 -----STSTEMTIVT-GTNGO-----PTDETIVIRTPTEGLITTTT 505
QY 304 QPPVKRLRLRGDSDGPPRARESERERDEQSPNVSLMQRMSDMLSRWFEASEVAQSN 363
Db 506 EP-----WIGFTSTSTEVTTITGTNGOPT-----DETIVIRTP 540
QY 364 RGRG-----RSRPRGGTQSQDIDLTPVPSPDLEVSETAMEVDTPAEQFL-----Op--- 411
Db 541 TSEGLITTTTTEPTWTFGFTSTSTEMTIVTGTNGOPTDETIVIRTPTEGLITRTTEPTWIG 600
QY 412 ----STSTMSAQAHSTSTSPESHSTPLSSPDSEORQVEASGHHTHQSDNNNEKLSP 468
Db 601 TPTSTSTEVTTITGTNGOPTD--ETVIVIRTPTAISSLSLSSSGQITSSITSSRPITP 658
QY 469 -KPGTGEPVLSLHYSTEGTTSTI-KLNFDEWSSSTASSRGGIGSHCKSEGQEESEFPPOS 526
Db 659 FYPNGTSSVSSSVTSSSLVTSSSFIS--SSVSSSTSTSTSFSESTSSVPTS 716
QY 527 SYQPPEGSETKAPESSEDVTKYQEGVSAENPVENH-----INITQSDKFTAKPL----- 577
Db 717 S--STSGSSEKTSASSSSSSSSSSESPPKSPNTNSSLPPVTSATTCQETASSLPAT 774
QY 578 DNSGERNDLNRGCGVPEESASSEKAKEPETSDOTSTESATNENNTNPEPOFOTEATG 637
Db 775 TTKTSEQTTLVTVTS---ESHVCTESISSAIVSTATVTVSGVTTTYTTCPISTETTK 831
QY 638 PSAHEETSTRDSALQDTSDDDPVL 663
Db 832 QTKGTQKGTQETQTTETTKQTTVV 857

Search completed: March 18, 2003, 15:40:34
Job time : 64 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:33:14 ; Search time 19 Seconds
(without alignments)
1331.774 Million cell updates/sec

Title: US-09-781-693A-2

Perfect score: 4527

Sequence:

Scoring table: 1 MSRGSGYPHLLNDVKKRGLG.....RLEGDRSGQENENEDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429.5	9.5	816	1	US-08-190-802A-54
2	429.5	9.5	816	4	US-08-477-346-54
3	429.5	9.5	816	4	US-08-473-089-54
4	429.5	9.5	816	4	US-08-487-072A-54
5	203	4.5	2137	4	US-09-134-001C-4463
6	192.5	4.3	985	5	PCT-US96-03916-6
7	192.5	4.3	985	5	PCT-US96-03916-66
8	189	4.2	1142	3	US-08-993-118-7
9	189	4.2	1142	3	US-08-845-528C-7
10	189	4.2	1142	4	US-08-061-709-2
11	188	4.2	1702	5	PCT-US95-10661A-5
12	188	4.2	1702	5	PCT-US95-10661A-5
13	185.5	4.1	571	4	US-08-961-083-4
14	175.5	3.9	1848	4	US-08-296-791-6
15	175.5	3.9	1848	5	PCT-US95-10661A-6
16	171.5	3.8	1964	2	US-08-790-912-3
17	171.5	3.8	2052	2	US-08-790-912-2
18	169	3.7	664	1	US-08-421-661-6
19	165.5	3.7	1187	1	US-08-320-559-28
20	165.5	3.7	1187	3	PCT-US94-04496-28
21	165.5	3.7	1187	5	PCT-US94-04496-28
22	165.5	3.7	1210	1	US-08-320-559-26
23	165.5	3.7	1210	3	US-08-445-860D-26
24	165.5	3.7	1210	5	PCT-US94-04496-26
25	165.5	3.7	1435	2	US-08-568-459A-4
26	165.5	3.7	1435	2	US-08-487-826B-4
27	165.5	3.7	1435	4	US-09-210-288-4

164	3.6	278	4	US-08-961-083-94	Sequence 94, Appl
154	3.4	713	1	US-08-190-802A-63	Sequence 63, Appl
154	3.4	713	4	US-08-477-346-63	Sequence 63, Appl
154	3.4	713	4	US-08-473-089-63	Sequence 63, Appl
154	3.4	713	4	US-08-487-072A-63	Sequence 63, Appl
154	3.4	1093	5	PCT-US93-03077-1	Sequence 3, Appl
153.5	3.4	50	1	US-08-190-802A-202	Sequence 1, Appl
153.5	3.4	50	4	US-08-477-346-202	Sequence 202, App
153.5	3.4	50	4	US-08-473-089-202	Sequence 202, App
152.5	3.4	50	4	US-08-487-072A-202	Sequence 202, App
152.5	3.4	1780	1	US-08-769-309A-5	Sequence 5, Appl
152.5	3.4	1780	3	US-08-994-570-5	Sequence 5, Appl
150.5	3.3	2843	1	US-07-741-940-2	Sequence 2, Appl
150.5	3.3	2843	1	US-08-289-548A-2	Sequence 2, Appl
150.5	3.3	2843	1	US-08-452-654-2	Sequence 2, Appl
150.5	3.3	2843	2	US-08-370-235A-2	Sequence 2, Appl
149.5	3.3	2843	4	US-08-449-731-2	Sequence 2, Appl
147.5	3.3	1805	1	US-07-853-913-2	Sequence 2, Appl
147.5	3.3	2842	1	US-07-741-940-7	Sequence 7, Appl
147.5	3.3	2842	1	US-08-289-548A-7	Sequence 7, Appl
147.5	3.3	2842	1	US-08-452-654-7	Sequence 7, Appl
147.5	3.3	2843	1	US-08-449-731-7	Sequence 7, Appl
147.5	3.3	2843	1	US-08-452-655B-2	Sequence 7, Appl
147.5	3.3	2843	1	US-08-452-655B-7	Sequence 7, Appl
147.5	3.3	2843	3	US-08-450-582-2	Sequence 7, Appl
147.5	3.3	2843	3	US-08-450-582-7	Sequence 7, Appl
147.5	3.3	2973	2	US-08-821-355A-7	Sequence 7, Appl
147.5	3.3	2973	2	US-09-003-687A-7	Sequence 7, Appl
147.5	3.3	369	4	US-09-136-605-7	Sequence 7, Appl
147.5	3.2	651	1	US-08-061-376-5	Sequence 5, Appl
147.5	3.2	651	1	US-08-769-309A-17	Sequence 17, Appl
147.5	3.2	894	3	US-08-362-525-22	Sequence 22, Appl
147.5	3.2	894	3	US-08-971-692-15	Sequence 15, Appl
147.5	3.2	1360	4	US-09-393-565-2	Sequence 2, Appl
146	3.2	1537	1	US-08-325-728A-2	Sequence 2, Appl
145	3.2	933	3	US-08-421-868-2	Sequence 2, Appl
145	3.2	933	3	US-09-421-868-2	Sequence 2, Appl
143.5	3.2	688	3	US-09-141-047-8	Sequence 8, Appl
143.5	3.2	1317	3	US-09-083-521-7	Sequence 7, Appl
141.5	3.1	1162	2	US-08-557-139-2	Sequence 2, Appl
141.5	3.1	1162	2	US-08-728-323A-2	Sequence 2, Appl
139.5	3.1	1162	2	US-08-557-139-2	Sequence 2, Appl
139.5	3.1	1584	3	US-08-656-068-2	Sequence 2, Appl
139.5	3.1	1850	4	US-09-457-040B-27	Sequence 27, Appl
139	3.1	548	3	US-08-656-034-10	Sequence 10, Appl
139	3.1	754	4	US-08-760-615-4	Sequence 4, Appl
139	3.1	1863	2	US-09-214-564A-2	Sequence 2, Appl
139	3.1	1863	2	US-08-603-753D-2	Sequence 2, Appl
139	3.1	1863	4	US-09-099-753-2	Sequence 2, Appl
139	3.1	1863	4	US-08-986-106-2	Sequence 2, Appl
138.5	3.1	340	1	US-09-007-678B-49	Sequence 49, Appl
138.5	3.1	340	4	US-08-190-802A-42	Sequence 42, Appl
138.5	3.1	340	4	US-08-477-346-42	Sequence 42, Appl
138.5	3.1	340	4	US-08-473-089-42	Sequence 42, Appl
138.5	3.1	949	1	US-08-487-072A-42	Sequence 42, Appl
138.5	3.1	1247	4	US-09-404-627-4	Sequence 4, Appl
138	3.0	571	4	US-09-404-627-2	Sequence 2, Appl
138	3.0	1346	2	US-09-134-001C-3865	Sequence 3865, Ap
137.5	3.0	341	1	US-08-635-121-2	Sequence 2, Appl
137.5	3.0	341	4	US-08-190-802A-45	Sequence 45, Appl
137.5	3.0	341	4	US-08-477-346-45	Sequence 45, Appl
137.5	3.0	341	4	US-08-473-089-45	Sequence 45, Appl
137.5	3.0	491	4	US-08-487-072A-45	Sequence 45, Appl
136.5	3.0	565	4	US-09-134-001C-3510	Sequence 3510, Ap
136.5	3.0	674	3	US-08-961-083-218	Sequence 218, App
136.5	3.0	750	4	US-08-899-852A-1	Sequence 1, Appl
136	3.0	340	1	US-08-165-239A-4	Sequence 4, Appl
136	3.0	340	1	US-08-190-802A-38	Sequence 38, Appl
136	3.0	340	4	US-08-477-346-38	Sequence 38, Appl
136	3.0	340	4	US-08-473-089-38	Sequence 38, Appl

Sequence 38, Appl
Sequence 20, Appl
Sequence 20, Appl
Sequence 21, Appl
Sequence 21, Appl
Sequence 19, Appl
Sequence 22, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 24, Appl
Sequence 16, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-190-802A-54
; Sequence 54, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R. 33,875
; REGISTRATION NUMBER: 8600-0139
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0960
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN, Fig. 37
; US-08-190-802A-54

Query Match 9.5%; Score 429.5; DB 1; Length 816;
Best Local Similarity 18.7%; Pred. No. 2.6e-25;
Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

QY 8 PHLLMDV---RKRSLGLEDPRLSRSLRGRREFIQRKLEATLNVHDGCVNTICWNTDGTG 63
Db 377 PRPRNWLVSALRDLRGLG---SSGRFVYACARLQVRSLEHVFEGHSGCVNTVHFNOHG 434
QY 64 EYLGSDDTKLIVISNPSYRKVLTTIRSGHRANITSAKFLPCTNDKQIVSCSGDGVLYFT 123
Db 435 TLLASGDDLVVWDLKRSVLFNDSGKNNITLQAKFLPNCNDAILAMCGRQGVRYA 494
QY 124 NVEQDAETNRQCFCHYGVTEIMTVDPYFLSCGEDGVTFVWFDTRIKTSCTKEDCK 183
Db 495 QLSAVAGTMTKRLVKHGASHRLGLEPDSFRFLTSCGDAVVENIDLRQHPASK---- 550
QY 184 DDLI---NCRRAATSVAIACPPPIPYLAVGSDSSVYIYDRMLGTATGNAYAGRTTG 239
Db 551 --LLVIKDGDKVGLTVFVNPANVYQFVAGGQDFMRYIDQKIDENV-----NNG 600
QY 240 MVARFIPSHL--NKKSCRTVSLCYSEDGQELVSVSSYIYLPKDDTARELTPSAEE 297
Db 601 VLKFCFPHLLSSDYPAHITSLMYSDGTETILLASYNDEDIYF----- 643
QY 298 RRELQPPVRLRLRGDMSDTGFRARPERERERDGEQSPNYSMLQRMDSMLSRWFEAS 357
Db 644 ----- 643
QY 358 EVAOSNRGRGRSRRGTQSOSDITLTPVSPDLVSETAMEVDTPAQFLQPTSTSTTM 417
Db 644 ----- 643
QY 418 SAQAHSTSPSTPSTPLSLSSPDSEORQSVESAGHHTHQSDNNNEKLSKPKGTGPVL 477
Db 644 ----- 643
QY 478 SLHYSTEGTITIKLNTDEWSSIASSSRGISGSHCKSEGESEFVPSQSVQPEGDSET 537
Db 644 ----- 643
QY 538 KAPESESDVTYQEGVSAENFVNHINITOSDKFTAKPLDSNGERNDLDRSCGVE 597
Db 644 ----- 643
QY 598 ESASSEKAKEPSTQSTESATNENNTNPPQFQTEATGSAHEETSTRDALSQDTHDS 657
Db 644 ----- 643
QY 658 DDDPVLIPGARYRAGPDRSARAVRIQEFRRRKERKEMEELDTLNIIRPLVWYKGRH 717
Db 649 D-----GAQY-----AKR-----YKGRH 661
QY 718 NSRTMIKEANFWG--ANFVMSGDCGHIFIDWRHTAHLMLLEAD-NHVVNCLQHPDPDP 774
Db 662 NNST-VKGVYVYGRSEFVMSGDCGHIFWEKSSQIVQFLEADEGGTINCIDSHPLP 720
QY 775 ILASSGIDYDIKIWSPLESRIFNRKLA--DEVITRNLMLLE--TRNTITVPAS---FM 827
Db 721 VLASSGLDHEVKIWSPIAEP---SKLAGLKNVIKINKLKRNTLURHSLFNNSMLCEL 777
QY 828 LRLASLNH-----IRADLEGDRSGSGQENENEDE 859
Db 778 MSHVTQSNYGRSWRGIRINAGGGDFSDSSSSSBEETNOE 815

RESULT 2
US-08-477-346-54
; Sequence 54, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,346
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,072
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: unknown

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: MTIS MUSCULUS PROTEIN

[illegible]

us-09-781-693a-2.ra1

wed Mar 26 09:38:53 2003

TITLE OF INVENTION: WD-40 - Derived Peptides and Uses

TITLE OF INVENTION: Thereof

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster

STREET: 2000 Pennsylvania Avenue, NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1812

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,072A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 2550-0025.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 816 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN, Fig. 37

US-08-487-072A-54

Query Match 9.5%; Score 429.5; DB 4; Length 816;

Best Local Similarity 18.7%; Pred. No. 2.6e-25;

Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

Query Match 9.5%; Score 429.5; DB 4; Length 816;

Best Local Similarity 18.7%; Pred. No. 2.6e-25;

Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

QY 8 PHLWDV----RKRSLGLEDPSSRLSRVLRGRRFIQRKLEATLNVHDGCVNTICWNTDGT 63

DB 377 PRPRNNVLSALDRQLG--SSGRFVYECAGARLEFVQRSLHFEHSGCCVNTVHFNQHG 434

QY 64 EYILSGDDTKLVISNPSYRKVLTIIRSGHRANIFSAKFLPTINDKQIVSCSGDGVIFYT 123

DB 435 TLLASGDDTKLVISNPSYRKVLTIIRSGHRANIFSAKFLPTINDKQIVSCSGDGVIFYT 494

QY 124 NYEQDAETNRQOQFCHYGTYYEIMTVNPDPYFLSCGEDGTVRWFDTRIKTSCTKEDCK 183

DB 495 QLSAVAGTHMTKRLVKGHSGHRLGLEPDSFPRFLTSGEDAVVFNIDLRQAHPASK 550

QY 184 DDILI-----NCRRAATSAICPPPIPYLLAVGCGSDSVRIYDRMLGTRATGNAGRGTTG 239

DB 551 --LLVTKDGGKVLGYTVFNPANVYQFVAGGQDQPMRIYDQKKIDENV-----NNG 600

QY 240 MVAREFIPSHL--NNKSCRVTSLCYSDGQELIVSYSSDIYLLFPKDDTARELKTSPSAEE 297

DB 601 VLKFCFPHLLSSDYPAHITSLMYSYDGTETILASYNDEDIYF----- 643

QY 298 RRELRQPPVKRLRLRGDWSDTGPRAPRESERERDGRSPNVSLMQMSDMLSRWFEEAS 357

DB 644 ----- 643

QY 358 EVAQSNRGRGRSRPRGTSQSDISTLPTVSPSPDLEVSETAMEVDTPAEQFLQFSTSTM 417

DB 644 ----- 643

QY 418 SAQAHTSSPTESPHSTPLSSPDSEQRQSVASGHHTHQSDNNNEKLSPKPGTGPVL 477

DB 644 ----- 643

QY 478 SLRYSTEGTTTSTIKLNFTEWSSIASRSGIGSHCKSEGQESFPVQSSVQPPGDSGT 537

DB 644 ----- 643

QY 538 KAPESEEDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGSERNDLNDRCGVPE 597

DB 644 ----- 643

QY 598 ESASSEKAKEPETSQDTSTESATNENTNPEPQFQTEATGPSAHEBSTRDSALQTDSDS 657

DB 644 ----- 643

QY 644 ----- 643

QY 658 DDDVPLIPGARYRAGPCDRSARVARIQEFRRRKRKEMPELDTLIRPLVMYKGRH 717

DB 644 ----- 643

QY 718 NSRTMIKEANFWG--ANFVSGSDCGHIFWDRHTAHLMLLEAD-NHVVNCLQPHFPDP 774

DB 662 NNST-VKGVYFGPRSEFVMSGSDCGHIFWKSQCVQFLEADGGTINCIDSHPLP 720

QY 775 ILASGDDYDKIWSPLEESRIFNRKLA--DEVITRNLMLLE--TRNTITVVAS---FM 827

DB 721 VLASGLDHEVKNWSPINAEF--SKKLAGLKNWIKINKLRDNFTLRHTSLFNNSMLCEL 777

QY 828 LRMLASLNH-----IRADLRGDRSESGQGENENEDE 859

DB 778 NSHVTSQNSYGRWGRGIRINAGGDFSDSSSSSEETINQE 815

DB 778 NSHVTSQNSYGRWGRGIRINAGGDFSDSSSSSEETINQE 815

RESULT 4

US-08-487-072A-54

; Sequence 54, Application US/08487072A

; Patent No. 6423684

; GENERAL INFORMATION:

; APPLICANT: Mochly-Rosen, Daria

; APPLICANT: Ron, Dorit

QY 478 SLHYSTEGTSTIKLNFTDEWSSIASRRGIGSHCKSEGOEESFVPOSSVQPPGDSST 537
 Db 644 -----
 QY 538 KAPESSBDVTKYQEGYSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPE 597
 Db 644 -----
 QY 598 ESASSEKAKEPETSQDOTSTESATNENNTNPEPQTEATGSAHEETSTRDSALQDTDS 657
 Db 644 -----
 QY 658 DDDPVLIPGARYRAGPCDRBSAVARIQEFFRRKEREKEMBELOTLNRRPLVKWYKGRH 717
 Db 649 D-----GAQY-----AKR-----YKGRH 661
 QY 718 NSTMIKEANFWG--ANFVMSGDCGHIFLWDRHTAEHLMLLEAD-NHVVNCLOPHFPDP 774
 Db 662 NNST-VKGYFYGPRGEFVMSGDCGHIFWEKSSQIVQFLEADGGGTINCIDSHPLYP 720
 QY 775 ILASSGIDYDIKWSPLESRIFNRKLA--DEVITRNEMLLEP--TRNITVTPAS---FM 827
 Db 721 VLASSGIDHEVKIWSPIAEP---SKLAGLKNVINKLKRDNFTLRHTSLFNMSMLCFL 777
 QY 828 LRLASLNH-----IRADRLGDRSGSGQENENEDE 859
 Db 778 MSHVTQSNYGRSWRGLIRINAGGDFSDSSSSSEETNQE 815

RESULT 5

US-09-134-001C-4463
 ; Sequence 4463, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134.001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4463
 ; LENGTH: 2137
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4463

Query Match 4.5%; Score 203; DB 4; Length 2137;
 Best Local Similarity 25.5%; Pred. No. 8e-07;
 Matches 89; Conservative 52; Mismatches 192; Indels 16; Gaps 9;

QY 317 SDTGPRARPERERDEGEQSPNVSLMQRSDMLSRWFEEASEVAQNRGRSRPRGRTS 376
 Db 1248 SDSASTSVSDSE-SASTSISESLSTSVSDSTS---TSTDASTSTSDSTSESL 1303
 QY 377 QSDISTL---PIVPSGPDLEVFETAMEVDTPAEQFLQSTSTSTMSAQASTSPSPESPHS 433
 Db 1304 SESISTSVSDSTASTSDASTSTSESESDASTSLSGSTSLSL---DSTSTSDSAS 1360
 QY 434 TPLLSGPDSEQSQVFAEGHHTHHSDNNNEKLSKPKGTGEPVLSLHYSTEGTTSTIKL 493
 Db 1361 TS-TSESDSE-RASVSLGGSTSTSLDSTSTSTSDASTSTSV-SDSNSASTSLGSLST 1417
 QY 494 NPTDEWSSIASRRGIGSHCKSEGOEESFVPOSSVQPPGDSKETAPRESSDVTKYQEG 553
 Db 1418 SVSDSTSTSDSAS-ASTSESDSERASTSLSGSTSTSLDSTSTSDASTSTSVSES 1476
 QY 554 VSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPEESASS--EKAKEPETS 611

Db 1477 NSTSISSESLSTVSDSTSTSTSDASTSTSVSDSASTSSSESVSTSDSESTSTSS 1536
 QY 612 DOTSTESTNNTNNTPEPQTEATGSAHEETSTRDSALQDTDDDD 660
 Db 1537 DSASTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASTSSSESD 1585

RESULT 6

PCT-US96-03916-6
 ; Sequence 6, Application PC/TUS9603916
 ; GENERAL INFORMATION:
 ; APPLICANT: Wild, Martha A.
 ; APPLICANT: Cochran, Mark D.
 ; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
 ; NUMBER OF SEQUENCES: 72
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/03916
 ; FILING DATE: 23-MAR-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/126,597
 ; FILING DATE: 24-SEP-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 39116-A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 985 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 PCT-US96-03916-6

Query Match 4.3%; Score 192.5; DB 5; Length 985;
 Best Local Similarity 20.7%; Pred. No. 1.6e-06;
 Matches 173; Conservative 101; Mismatches 290; Indels 273; Gaps 41;

QY 172 RIKTSCTKEDCKDILLINCRRATSAICPPIPYLAGVSDSSSVIYDR-RMIGTRATG 230
 Db 146 RLETPCDENFLQNEPTWGSKR---WLGPPSPY-----VRDNDVAVLTAKAIGECYSN 195
 QY 231 NYAGRTGTGVARFI--PSHLNN-----KSCRVTSLCYSDGGQELVSYSSDYILF 280
 Db 196 SAAQTGLTSLNMTFFYPKPRIVNTWTGGPSPRITVYSSRENGQPVLRNVSDGLVKY 255
 QY 281 DKPDD-----TA-RELKTPSAERREELQPPVKRLRLRGDWS 317
 Db 256 TPDIDCRAMINVIANYSPADSGSVLAFTAFRCGLPSAQLH-----RIDMS 302
 QY 318 DTGPRARPERERDEGEQSPNVSLMQRSDMLSRWFEEASEVAQNRGRSRPRGRTSQ 377
 Db 303 GTEP---PGTETTFD-----CQKM-----IETPYRALGSNVPRDDSTR 337
 QY 378 SDISTLTPVSPDLE--VSETAMEVDTPAEQFLQSTSTSTMS--AQASTSSPTESP-- 431

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

```

338 PGATLPPEDTAAPDFDCTSTPTTVPPEAITLIPRSTSDMGFFSTARATGSETLSVPV 397
432 -----HSTPL---LSSPDSEQ-----RQSVASGHHTHSDNNNEKLSKPG 471
398 QETDRTLTSTLPLTPGESENLFPTTAPGISTETPSAAHETQTQSAETVFTQSPS 457
472 T-----CEPVLHSHYSTEGTTSIKLNFTD--EWSSIASSSRGI-----GS 511
458 TESEARQSQEP-----WYFTQTPSTEQAALTOQIAETEAFTQTPSAEQMTFTQPGA 513
512 HCKSEGOESFVP-----QSSVQPPPE-GDSETKAPE---ESSEDVTKYQEGVSAENPVNH 563
514 ETEAPAQTPSTPEIFTOQSRSTPPTARAPSAAPVFTQSSSTVTEV----- 560
564 INTQSKDTAKPLDSNGERNDLNRSCGVPEESASSEKAKEPETSDQISTES----- 618
561 --FTQTPSTVPKTLSSSTPEPAIFTRTQSAET--EAFQTSSAEPDTRMTQSTETHEFTQ 616
619 -----ATNENNTNPEPQFTEAT-----GPSAHEETSTRDSA----- 650
617 APSTVPKATQTPSTEPEVLTQSPSTPEVPTFTLGAEPETITQTPSAPEVYTRSSSTMP 676
651 -LQTDSDDDPVLIPGARYRAGPDRRS-----AVARIQEFFRRRKR-----KE 695
677 TAQSTPLASQNP-----SSGTGTHNTEPRTPYVQTPHTQKLYTENKTLSPFTVWSE 729
696 MEELDTLNRPL--VKMVKYKGRN-----SRMTK-----EANFWGANFV-----MS 736
730 FHEMSTAESQTPLLDKVIVEKFSNDGEVATCVSTVKSPYRVTN--WKVDLDVDMDEIS 788
737 GSDCGHIF-----IWDRTAHLMLLEADNHVNCVLCQPHP-----FDPILASSG 780
789 GNSPAGVFNSEKWKQKLYRYVTDGRTSVQLMCLSCISH-----SPEPYCLDTSLIAR- 842
781 IDYDIKWSPLEESRIFNRKLADEVITRNELMLEETRTITVPASFMLRLASLNHI 837
843 -----EKDIAPELYFTSD--POTAYCTITLPSGVVPRFEWSLNNV 880

```

```

RESULT 7
PCT-US96-03916-66
; Sequence 66, Application PC/TUS9603916
; GENERAL INFORMATION:
; APPLICANT: Wild, Martha A.
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESS: Cooper & Dunham LLP
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/03916
; FILING DATE: 23-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,597
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39116-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400

```

```

TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 985 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-03916-66

```

```

Query Match 4.38; Score 192.5; DB 5; Length 985;
Best Local Similarity 20.7%; Pred. No. 1.6e-06;
Matches 173; Conservative 101; Mismatches 290; Indels 273; Gaps 41;

```

```

QY 172 RIKTSCTKEDCKODILINCRRATSVAIICPIPIYLAAGCSDSSVRIYDR-RMLCTRATG 230
Db 146 RLETPCDENLQNEPTWGSKR-----WLGPPSPV-----VRDNDVAVLTKAQYIGECYSN 195
QY 231 NYAGRTGTGMVARI--PSHLNN-----KSCRVTSLCYSEGDQEBILYSYSDIYLF 280
Db 196 SAAQTGLTSLNMTFFYFKRIVNVTWTGGPSPSRIVVYSRENGQPVLRNVSDGFLVKY 255
QY 281 DPKDD-----TA-RELKTPSAEERREELRPPVKRLRLGDWS 317
Db 256 TPDIIDGRAMINVIANSYSPADSGSVLAFATFREGKLPALQLH-----RIDMS 302
QY 318 DTGPRARPESEERERDGEQSPNVSLMORMSMLSRWFEEASEVAQSNRGRSRPRGTSQ 377
Db 303 GTEP---PGTETTFD-----COKM-----IETPYRAGLSNVPRDDSI 337
QY 378 SDISLTPVPPSPDLE--VSETAMEVDTPAEQFLOPSTSSMS--AQAKSTSSSTESP-- 431
Db 338 PGATLPPEDTAAPDFDCTSTPTTVPPEAITLIPRSTSDMGFFSTARATGSETLSVPV 397
QY 432 -----HSTPL---LSSPDSEQ-----RQSVASGHHTHSDNNNEKLSKPG 471
Db 398 QETDRTLTSTLPLTPGESENLFPTTAPGISTETPSAAHETQTQSAETVFTQSPS 457
QY 472 T-----CEPVLHSHYSTEGTTSIKLNFTD--EWSSIASSSRGI-----GS 511
Db 458 TESEARQSQEP-----WYFTQTPSTEQAALTOQIAETEAFTQTPSAEQMTFTQPGA 513
QY 512 HCKSEGOESFVP-----QSSVQPPPE-GDSETKAPE---ESSEDVTKYQEGVSAENPVNH 563
Db 514 ETEAPAQTPSTPEIFTOQSRSTPPTARAPSAAPVFTQSSSTVTEV----- 560
QY 564 INTQSKDTAKPLDSNGERNDLNRSCGVPEESASSEKAKEPETSDQISTES----- 618
Db 561 --FTQTPSTVPKTLSSSTPEPAIFTRTQSAET--EAFQTSSAEPDTRMTQSTETHEFTQ 616
QY 619 -----ATNENNTNPEPQFTEAT-----GPSAHEETSTRDSA----- 650
Db 617 APSTVPKATQTPSTEPEVLTQSPSTPEVPTFTLGAEPETITQTPSAPEVYTRSSSTMP 676
QY 651 -LQTDSDDDPVLIPGARYRAGPDRRS-----AVARIQEFFRRRKR-----KE 695
Db 677 TAQSTPLASQNP-----SSGTGTHNTEPRTPYVQTPHTQKLYTENKTLSPFTVWSE 729
QY 696 MEELDTLNRPL--VKMVKYKGRN-----SRMTK-----EANFWGANFV-----MS 736
Db 730 FHEMSTAESQTPLLDKVIVEKFSNDGEVATCVSTVKSPYRVTN--WKVDLDVDMDEIS 788
QY 737 GSDCGHIF-----IWDRTAHLMLLEADNHVNCVLCQPHP-----FDPILASSG 780
Db 789 GNSPAGVFNSEKWKQKLYRYVTDGRTSVQLMCLSCISH-----SPEPYCLDTSLIAR- 842
QY 781 IDYDIKWSPLEESRIFNRKLADEVITRNELMLEETRTITVPASFMLRLASLNHI 837
Db 843 -----EKDIAPELYFTSD--POTAYCTITLPSGVVPRFEWSLNNV 880

```

```

RESULT 8
US-08-993-118-7
; Sequence 7, Application US/08993118

```

Patent No. 5997872
 GENERAL INFORMATION:
 APPLICANT: LUCAS, Sophie;
 APPLICANT: DE SMET, Charles;
 APPLICANT: BOON-FALLEUR, Thierry
 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
 TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Felfe & Lynch
 STREET: 805 Third Avenue
 CITY: New York City
 STATE: New York
 COUNTRY: USA
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
 COMPUTER: IBM PS/2
 OPERATING SYSTEM: PC-DOS
 SOFTWARE: Wordperfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/993,118
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/845,528
 FILING DATE: April 25, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Mary Anne Schofield
 REGISTRATION NUMBER: 36,669
 REFERENCE/DOCKET NUMBER: LUD 5455
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 688-9200
 TELEFAX: (212) 838-3884
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1142
 TYPE: amino acids
 STRANDEDNESS: single-stranded
 TOPOLOGY: linear
 US-08-993-118-7

Query Match 4.2%; Score 189; DB 2; Length 1142;
 Best Local Similarity 21.0%; Pred. No. 3.9e-06;
 Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;
 QY 271 SYSDYIYLPDKDDTARE-----LKTPSAERERREELRQPPVKRLRLRGDWSDT-G 320
 DB 486 SSSSTLLSLFQSSSPECTQSTFEGFPQSLQIPQSPPEGENTHS-PLQIVPSLPPEWEDSLS 544
 QY 321 PRAPESERERDGEQSPNV---SLMQMSDMLSRWFEEA--SEVAQSNRGRGRSRPRGGT 375
 DB 545 PHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGED 604
 QY 376 SOSDITLTVPSPDLEVSETAMEVDTPAQFLQSTSTMSAQAHSTSSPSPHSTP 435
 DB 605 SMSPL-----YFPQSP-LQGEFQSSLSQSPVSCSTPSSLPQSPPEQSPQSP 659
 QY 436 LLSSPDSEQRQSVASGHHTTHQSDNNNEKLSL-----KPGTGEPLV-SLHY--STEGTT 487
 DB 660 -LHSPQSPPEGMHSPQSPQSPESAPAGEDSLSPQIPQSPLEGEDSLSLHPPQPPWE 718
 QY 488 TSTIKLNT-----DEWSSIASSTRGIGSHCKSGQESFPQSVQSPPEGDSLT--KA 539
 DB 719 DSLSPHPFPQPPQGEDPQSSLSQSPVSCSTSLSLQSPF-PESQSPPEGPAQSPQR 777
 QY 540 P-----EESDVTYKQEGVS-----AENPVENHINITQSDKFTAKPLDLSNCE 583
 DB 778 PVSFFSTLASLQSHSPQSPPEGPAQSPQSPVSSPSSSSLSQSPVSS-----833
 QY 584 RNDLNRSCGVPESASSEKAKPEPESDQSTESATNNTNPNPEPQTATGPSAHEE 643

DB 834 -----FPSSTSSLSKSSPESPLQSPVIFSSTSLSP-----FSESSSP-VDEY 878
 QY 644 TSTRDSALQDQDSDDDPVLIPGARAGPGRDRGSAVAR-----682
 DB 879 TSSDITLESLSLTDSESLIESEPLFTVTLDEKVDLAFRLLLKYQVQKPTKAEMLTNV 938
 QY 683 -----IQEFFRRRKRKE-----MEELD-----TLNIRPLVKMYKGRHNSR 720
 DB 939 ISRTGYFVFIRKAREFIELFGISLREVDPPDSYFVNTIDLTSEGCLSDQSGMSQNR 998
 QY 721 TMIKEANFWANFVMSGDCGHFIWD-----RHTAEHLMLLEADNHNVC-LQPH-- 770
 DB 999 LLILILSI-----IFIKTYAEEVINDVLSGIGVYRAGREHFAFGEPRELLTKVWQEHYL 1054
 QY 771 PFDPIASSGIDYDIKIWSPLEESRIFNRKLADDEVITRNELMLEETRT--ITVPASF 836
 DB 1055 EYREVPNSPPRYEF-LWGPRAHSEVIRKVV-----FLAMLKNTVPIITFPSSY 1103
 RESULT 9
 US-08-845-528c-7
 ; Sequence 7, Application US/088455528c
 ; Patent No. 6027924
 ; GENERAL INFORMATION:
 ; APPLICANT: LUCAS, Sophie;
 ; APPLICANT: DE SMET, Charles;
 ; APPLICANT: BOON-FALLEUR, Thierry
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR
 ; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGE-C1 AND USES
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Felfe & Lynch
 ; STREET: 805 Third Avenue
 ; CITY: New York City
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
 ; COMPUTER: IBM PS/2
 ; OPERATING SYSTEM: PC-DOS
 ; SOFTWARE: Wordperfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/845,528c
 ; FILING DATE: April 25, 1997
 ; CLASSIFICATION: 4335
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mary Anne Schofield
 ; REGISTRATION NUMBER: 36,669
 ; REFERENCE/DOCKET NUMBER: LUD 5455
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 688-9200
 ; TELEFAX: (212) 838-3884
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1142
 ; TYPE: amino acids
 ; STRANDEDNESS: single-stranded
 ; TOPOLOGY: linear
 ; US-08-845-528C-7
 Query Match 4.2%; Score 189; DB 3; Length 1142;
 Best Local Similarity 21.0%; Pred. No. 3.9e-06;
 Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;
 QY 271 SYSDYIYLPDKDDTARE-----LKTPSAERERREELRQPPVKRLRLRGDWSDT-G 320
 DB 486 SSSSTLLSLFQSSSPECTQSTFEGFPQSLQIPQSPPEGENTHS-PLQIVPSLPPEWEDSLS 544
 QY 321 PRAPESERERDGEQSPNV---SLMQMSDMLSRWFEEA--SEVAQSNRGRGRSRPRGGT 375
 DB 545 PHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGED 604

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

```

376 QY SQSDISTLPTVPSSPDLEVSAMEVDTPAEQFLOPSTSTMSQAQHSSTSPESHPTP 435
376 QY SQSDISTLPTVPSSPDLEVSAMEVDTPAEQFLOPSTSTMSQAQHSSTSPESHPTP 435
605 DB SMSPL-----YFPQSP-LQGEFQSLQSPVSCSSSTPSSLPQSPFESSQSPGPGVQSP 659
605 DB SMSPL-----YFPQSP-LQGEFQSLQSPVSCSSSTPSSLPQSPFESSQSPGPGVQSP 659
436 QY LLSPPDSEORQSVASGHHHTHQSDNNNEKLS-----KPGTGPVL-SLHY--STEGTT 487
436 QY LLSPPDSEORQSVASGHHHTHQSDNNNEKLS-----KPGTGPVL-SLHY--STEGTT 487
660 DB -LHSPQSPGEGHMQSPLOSPESAPAGEDSLPLQIPQSPLEGEDSLSLHFPQSPPEWE 718
660 DB -LHSPQSPGEGHMQSPLOSPESAPAGEDSLPLQIPQSPLEGEDSLSLHFPQSPPEWE 718
488 QY TSTIKLNT-----DEWSSIASRRGIGSHCKSEGOEESFVQSPVQPPGDSST--KA 539
488 QY TSTIKLNT-----DEWSSIASRRGIGSHCKSEGOEESFVQSPVQPPGDSST--KA 539
719 DB DLSPLHFPQPPQGGEDFQSLQSPVSCSSSTSLQSPF-PESQSPGPGPAQSPLOQR 777
719 DB DLSPLHFPQPPQGGEDFQSLQSPVSCSSSTSLQSPF-PESQSPGPGPAQSPLOQR 777
540 QY P-----EESSEDVTYKQEGVS---AENPVENHINITQSDKFTAKPLDSSNGE 583
540 QY P-----EESSEDVTYKQEGVS---AENPVENHINITQSDKFTAKPLDSSNGE 583
778 DB PVSSFFSYTLASLQSSHESPPQSPGPAQSPLOSPVSSFPSSSTSSLSQSPVSS----- 833
778 DB PVSSFFSYTLASLQSSHESPPQSPGPAQSPLOSPVSSFPSSSTSSLSQSPVSS----- 833
584 QY RNDNLDRSCGVPESASKEKAPETSDQSTESATNENNTNPEFQFOTATGSAHEE 643
584 QY RNDNLDRSCGVPESASKEKAPETSDQSTESATNENNTNPEFQFOTATGSAHEE 643
834 DB -----FPSSTSSLSKSSPESPLQSPVIFSSSTSLSP---FSESSSP-VDEY 878
834 DB -----FPSSTSSLSKSSPESPLQSPVIFSSSTSLSP---FSESSSP-VDEY 878
644 QY TSTRSALOQTDSDDDPVLPGARYRAGGDRRSARV----- 682
644 QY TSTRSALOQTDSDDDPVLPGARYRAGGDRRSARV----- 682
879 DB TSSDYLLESLSLDSLSLSEPLTYTLDEKVDLAREFLLLKYQVKQIPITKAEMLTNV 938
879 DB TSSDYLLESLSLDSLSLSEPLTYTLDEKVDLAREFLLLKYQVKQIPITKAEMLTNV 938
683 QY -----IQEFFRRKERKE-----MEELD-----TLNIRRLPLVMYKGRHSR 720
683 QY -----IQEFFRRKERKE-----MEELD-----TLNIRRLPLVMYKGRHSR 720
939 DB ISRYTGVPVIFRKAREFIEILFGISLREVDPDSDSVVFNLTDLTSEGCLSDGQMSQNR 998
939 DB ISRYTGVPVIFRKAREFIEILFGISLREVDPDSDSVVFNLTDLTSEGCLSDGQMSQNR 998
721 QY TMKEANFWANFVMSGDCGHIFWD-----RHTAHLMLLEADNHVNC-LQPH-- 770
721 QY TMKEANFWANFVMSGDCGHIFWD-----RHTAHLMLLEADNHVNC-LQPH-- 770
999 DB LLILLSI-----IFKTYASEEVIWDLGIGVRAGREHFAFGEPRELLTKVWQEHYL 1054
999 DB LLILLSI-----IFKTYASEEVIWDLGIGVRAGREHFAFGEPRELLTKVWQEHYL 1054
771 QY PFDPIASSGIDYDIKINSPLESRIENRKLADDEVITRNELMLEETNT--ITVPASF 826
771 QY PFDPIASSGIDYDIKINSPLESRIENRKLADDEVITRNELMLEETNT--ITVPASF 826
1055 DB EYREVPNSPPRYEF-LMGPRAHSEVIRKKVVE-----FLAMKNTVPTITPSSY 1103
1055 DB EYREVPNSPPRYEF-LMGPRAHSEVIRKKVVE-----FLAMKNTVPTITPSSY 1103

RESULT 11
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299

RESULT 10
US-09-061-709-2
; Sequence 2, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gurs, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 2
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-061-709-2

Query Match 4.2%; Score 189; DB 4; Length 1142;
Best Local Similarity 21.0%; Pred. No. 3.9e-06;
Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;

QY 271 SYSSDYILPDKDPTARE-----LKTSAERREERLEPPVYKRLRLGDWSDT-G 320
DB 486 SSSSTLLSLOFQSSPECTQSFEGFPQSPLOIPQSPGENTHS-PLQIVPSLPEWEDSL 544
QY 321 PRAPERERNDGQSPNV---SLMQMSDMLSEWFEA---SEVAQSNRGRSRPRGGT 375

```


us-09-781-693a-2.rai

wed Mar 26 09:38:53 2003

Db 110 SDSQEQAEHKKLEKKEKISPKETGVTNLNPODEVLSQOLNKPPELLYREETHETID 169
QY 272 Y8S-----DYIYLFDPKDDTARELKTSAERREELRQP 305
Db 170 FOEIEQENPLAAGTVVRVQKQGLKVKKIVRIFSVNKEEVSREIVSTST--TA 221
QY 306 PVKRLRLRGDMSDTGPRAPESERERDGEQS-----PNVSLMQMSDMLSRWFEEASEVA 360
Db 222 PSPIRVEKGGKVKQVKEQPEETGVHEKDVQSGAIVRPAIQ-----PELPEAV 268
QY 361 QSNRGRGRSRPRGTSOSDITLPTV-----PSSPDLEVSETA----- 398
Db 269 VSDKGEVOP-----TLPEAVVTGKETEVOPESDPTVWSDKGEQVAPLPEY 318
QY 399 -----NEVDTPAQFQPTSTSTMSAQASTSPSTPHSTPLSSPDSEFORQSVASG 452
Db 319 KGNTEQVKPEPVEKTEQGEPEE-----PVKPTETPVNPNNEGTEGTSIQEAE 371
QY 453 HHTH--HOSDNNNEKLSKPGTGEVLSHYSTEGTTSTIKLFTDEWSIASSSRGIG 510
Db 372 NPVQPAEESTTNSEKVP-----DTSKNTGEVSSNPSTSTSVGESNKEPH 418
QY 511 SHCKSEGOEESFVQSSVQPPGDSSETKAPESSESDVTKYQGVSAENPVENHINITQSD 570
Db 419 NDSKNENSEKT-VEEVVNPNEGTVEGTSNQETKVPQPAEE-----TOTN 463
QY 571 KFTAKPLDSNGERDNLDRSCGVPEESASEK---AKEPETSDOTSTESATNENNTNP 627
Db 464 --SGKIANENTGEVSNKPSDSRPPV-EESNQPEKNGTATKPNESGNTTSEN---GOTEP 516
QY 628 EPQFQTEATGPSAHEETSTRDSALODTDDDDDP 661
Db 517 EPSNGNSTEDVSTESNTSNGNEEKQENELDP 550

RESULT 14
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids

QY 459 SDNNNEKLSKPGTGEVLSHYSTEGTTSTIKLFTDEWSIASSSRGIGSHCKSEGO 518
Db 1212 TENT-----TDQP-----TEREKTAKVETETQPPQVASQA-----SPKQEQ 1249
QY 519 EESFVQSSVQPPGDSSETKAPESSESDVTKYQGVSAENPV-ENHIN-----ITQSDK 571
Db 1250 SETVOQAVLESENVPVNNAAEVOAQLOTQTSATVSTKQAPENSINTGSATITAE 1309
QY 572 FYAKPLDSNGERDNLDRSCGVPEESASEKAKEPETSDQSTESATNENNTNPEQF 631
Db 1310 KSDKP-----QTETAASDASQKANTVADNSVANSESSEPKSRR 1351
QY 632 QTEATGPSAHEETSTRDSALODTDDDDVLPGARVYRAGDGRSARVARIOEFRERK 691
Db 1352 RRSISQ---QETSABETTAASDTET-----TIADNSKRKSKP-NRRS 1392
QY 692 ERKEMEELDTLIRRLVVKWYKGRHSRTMIKEANFWANFGVMSGDCGHIFI---WDR 748
Db 1393 VRSE-----PTVYNGSDRSVALRDLTSTNTNAVISDAMAKAQFVALNVCK 1438
QY 749 HTAEHLMLLEADN 761
Db 1439 AVSQHISOLENN 1451

RESULT 13
US-08-961-083-4
; Sequence 4, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Hp Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-4
; 4.1a: Score 185.5; DB 4; Length 571;
Query Match
Best local similarity 19.3%; Pred. No. 2.6e-06;
Matches 99; Conservative 80; Mismatches 198; Indels 137; Gaps 17;
QY 212 SDSVRIYDRMLGTRATGNVAGRGTTGTVARFIPSHLNKSCRVTSLCYSDGQELVLS 271

```

RESULT 15
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; ADDRESS CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:

us-09-781-693a-2.rai

Wed Mar 26 09:38:53 2003

649 SNOPEKNGTATKPSNGTISENGQTEPEKLELRNVSDIELYSQTNGT-----YRQHVS 704

648 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/790,912

FILING DATE: 29-JAN-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/026,939

FILING DATE: 23-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Leary, Kathryn

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 7600-401

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 567-2020

TELEFAX: (215) 567-2991

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1964 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-790-912-3

Query Match 3.8%; Score 171.5; DB 2; Length 1964;

Best Local Similarity 18.1%; Pred. No. 0.00023;

Matches 121; Conservative 101; Mismatches 229; Indels 217; Gaps 24;

111 IVSCSGDGVIFVTNVEQDAETNROCQFCHYGTYYEI-----MTVPND--PYTFLSGCGD 163

139 LVAGMNGVLVSVHALEN-----HLLNNTDYELTSGEKLPLPKKEISGYTIVGIKE 191

164 G-----TVWFDTRIKTSCTK----- 179

192 GKTTDFEVSNOEKSAAFTKQKVDYNTVNFVDPHTVQAIQETPVSTKTEVQVV 251

180 EDCKDDILINCRRAATVAICPPPIYLLAVGSDSSVRIYDRMLGTRATGNAGRGTTG 239

252 EKPFSTELINPRKEEQ-----SSDSQEQLAHKNLETKKEEKISPKKKTG 297

240 MVAFIPSHLNKSCRVTSICYSGEDGOEILVSYSS----- 274

298 VNTLNPODEVLSGQLNKPELLYREEIETFKIDFQEEIQENPDLAECTVRVQEGKLGKV 357

275 DYIYLFDPKDDTARELKTPTSAERREELQPPVKRLRLGDSWDGPRARPESERERDG 333

358 ELVRIFSVNKEEVSREIVST-----TAPSPRIVEKGTGKTQVKEQPTGVYHKD 409

334 EQS-----PNVSLMQNSDMLSRWFEBASEVAQSNRGRSRPRGGTSQSDISTLPTV-- 386

410 VQSGAIVEPAIQ-----PELPEAVVSDKGEPEVQP-----TLPEAVV 446

387 -----PSSPDLEVSGETA-----MEVDTPEAQFLOPSTSSYMSAQ 420

447 TDKGETEVQESPDTVWSDKGEQVAPLRYKGNIEQVKPETVPEKKEQGEKTEEV- 505

421 AHSTSSPTSPHSHTPLSSPDSEQRQSVASGHHTH--HQSDNNNEKLSRPGTGEVPLS 478

506 -----PVKPEETPVNNEGTTGTSIQEAENPVQPAEESTNSEKVP----- 549

479 LHVSTEGTTTSLKLTDEWSIASRGICSHCKSGQESFVPQSSVQPPGDSSTK 538

550 ----DTSSSENTGEVSSNPSTTSVGSNKPENHDSKNSEKT--VEEVPVNPNEGTVG 605

539 APESSEEDVTKYQGVSAENPVENHINITQSDKFTAKPLDLSNGSGERNDLNRSCGVPE 598

606 SNQETKPVQPAEE-----TQTN--SGKIANENTGEVSNKPSDKPPV-EE 648

599 SASSEK---AKEPETSQDTSTESATNE-----NTNPEPQFQTEATGSAHEETSTR 647

649 SNOPEKNGTATKPSNGTISENGQTEPEKLELRNVSDIELYSQTNGT-----YRQHVS 704

648 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

649 DSALQDTD 655

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-320-559-28

Query Match          3.7%; Score 165.5; DB 1; Length 1187;
Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LFDPRDARELKTSAERREELRQPPVKRLRGDNDSTGP-----RARPSEERD 332
DB 240 LVAPAPQSPQTFPPLSLPSKSVAMQKPTAYVRPM-DGQDQAPSPKLPDYROOT 298
QY 333 GEQS-----PNVSLQRMS-----DMLSRW----- 352
DB 299 FEKTLKVPKAKLTKLKPMSQVETGYNEVHCVEELKEMTHSWPPLTAIHTPSTAE 358
QY 353 -----FEAEVAQSNRGRSRPRGGTQSQSDISTLTPTVPSSDPLEVSEAMEVD 402
DB 359 PSKEPPTKDSQVSSVTQKQYDTSKTHSNSQOQTSSM-----LEDDLQSDSE--D 411
QY 403 TPAPQFLQ-PSTSTMSAQAHSTSTESPHSTPLSSPDSEQRQVEASGHHTHQSDN 461
DB 412 SDSQTEPKPPSSAPPAPQSLPEVASAHS-----SSAESESTSDSS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSHYSTEGTTTSTIKLNFDTGW-----SSIASSSRGIGS----- 511
DB 461 ESESSSDSENELETPAPEPEPTTNKWL-----DNWLTQVSPAPAPPGPKRSTPPRR 517
QY 512 HCKSEGQESVQ--SSVQPPGDSSEKAPESSE-----DVTKYQGVSAENPVENHN 565
DB 518 HPESKSSDSATSQHSSEKDPFPKSSSKAPAPPEAPHGPKRSCQKSPAQEPQ-- 575
QY 566 IQSDKFTAKPL--DSNNGERNDLMDRCV-----PEESASS 602
DB 576 -TVGTYQPKPKPVKASARAGSRSTSLQGEREPGLLPYGRDQTSKDKPKVTKGRPRAASN 634
QY 603 E-KAKEPETSQGTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
DB 635 EPKPAVPPSSSEKKHKSLSLPAPSKALSPEKADKVEDRTPHFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLP----- 665
DB 694 SGRSTSGCQAVVQEDSRKRLPLPLRDTKLPLSDPLRTPPQSLMVKITDILLSRIQOP 753
QY 666 ---GARYRAG-----PGRRSNAVRIQEFRR-RKREKEMEELDTLNIRREPLVKMYKG 715
DB 754 PGKGRQRKAEDKOPAPGAKKHSEKSSSSSKLAKKRGAEKRCODNKKIRLEIKSQ 813

```

```

QY 716 HRNRTMIKEAN 727
DB 814 SSSSSSHKSS 825

RESULT 20
US-08-545-860D-28
; Sequence 28, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-545-860D-28

Query Match          3.7%; Score 165.5; DB 3; Length 1187;

```


Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LDPKDDTARELKTPSAEERREELRQPPVKRLRGDWSDTGP-----RARPESERERD 332
Db 240 LVAPAPPSQTFPPPSLPKSVAMQOKPTAYVRPM-DGQDQAPSESPKLPDPEDYRQQT 298
QY 333 GEQS-----PNVSLMORMS-----DMLSRW-----SDS 352
Db 299 FEKTDLVKPAKAKTLKMPFSQSVQEQYTSNEVHCVEILKEMTHSWPPPLTAHTPTSTAE 358
QY 353 -----FEASEVAQSNRGRGRPRGGTQSQDISLTPVPSPPDLEVSETAMEVD 402
Db 359 PSKFPPTKDSQHVSVYTNQKQYDTSKTHSNQOCTSSM-----LEDDQLQSDSE--D 411
QY 403 TPAEQFLQ-PSTSTMSAQAHSTSPSPHSTPLSSPDSEQRQSVASGHHTHQSDN 461
Db 412 SDEQPEKPPSSAPPSAPQSLPEPVAHNS-----SSAESTSDSDS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSLHYSTEGTTSTIKLNFTEW-----SSIASSSRGIGS----- 511
Db 461 ESESSSDSENEPLETPAPEPEPTTNKWL-----DNMLTKVSPAPAPPEGPRSTPEPR 517
QY 512 HCKSEQGESFVQ--SSVQPEGDSQETKAPESSE-----DVTKYQEGVSAENPVENHIN 565
Db 518 HPESKSSDSATSQHSESKDPPKSSKAPRAPPEAPHGKRSCQKSPAQOEPQRO-- 575
QY 566 ITQSDKFTAKPL--DSNNGERDNLDRSCGV-----PEESASS 602
Db 576 -TVGTPKPKPKVKASARAGRTSLQGEREGLLPYGRDQTSKDKPKVKTKGRPRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
Db 635 EPKPAVPPSEKKHKSLSLAPSKALSGPEAKDNVEDRTPHEFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLP----- 665
Db 694 SGRTSGRQAVVQEDSRKDRPLPLDRDKLLSPLRDTPPQSLMWKITLDLLSRIPQ 753
QY 666 ---GARYRAG-----PGDRRSARVARIQEFRR-RKEREMEELDTLNIRRLPVKMYKG 715
Db 754 PGKSRQKAEQKPPAGKHSSEKSSSSSLAKKRKGEARDCDNKKIRLEKEIKSQ 813
QY 716 HRNSRTMIKEAN 727
Db 814 SSSSSSSSHKSS 825

Query Match 3.7%; Score 165.5; DB 5; Length 1187;
Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LDPKDDTARELKTPSAEERREELRQPPVKRLRGDWSDTGP-----RARPESERERD 332
Db 240 LVAPAPPSQTFPPPSLPKSVAMQOKPTAYVRPM-DGQDQAPSESPKLPDPEDYRQQT 298
QY 333 GEQS-----PNVSLMORMS-----DMLSRW-----SDS 352
Db 299 FEKTDLVKPAKAKTLKMPFSQSVQEQYTSNEVHCVEILKEMTHSWPPPLTAHTPTSTAE 358
QY 353 -----FEASEVAQSNRGRGRPRGGTQSQDISLTPVPSPPDLEVSETAMEVD 402
Db 359 PSKFPPTKDSQHVSVYTNQKQYDTSKTHSNQOCTSSM-----LEDDQLQSDSE--D 411
QY 403 TPAEQFLQ-PSTSTMSAQAHSTSPSPHSTPLSSPDSEQRQSVASGHHTHQSDN 461
Db 412 SDEQPEKPPSSAPPSAPQSLPEPVAHNS-----SSAESTSDSDS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSLHYSTEGTTSTIKLNFTEW-----SSIASSSRGIGS----- 511
Db 461 ESESSSDSENEPLETPAPEPEPTTNKWL-----DNMLTKVSPAPAPPEGPRSTPEPR 517
QY 512 HCKSEQGESFVQ--SSVQPEGDSQETKAPESSE-----DVTKYQEGVSAENPVENHIN 565
Db 518 HPESKSSDSATSQHSESKDPPKSSKAPRAPPEAPHGKRSCQKSPAQOEPQRO-- 575
QY 566 ITQSDKFTAKPL--DSNNGERDNLDRSCGV-----PEESASS 602
Db 576 -TVGTPKPKPKVKASARAGRTSLQGEREGLLPYGRDQTSKDKPKVKTKGRPRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
Db 635 EPKPAVPPSEKKHKSLSLAPSKALSGPEAKDNVEDRTPHEFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLP----- 665
Db 694 SGRTSGRQAVVQEDSRKDRPLPLDRDKLLSPLRDTPPQSLMWKITLDLLSRIPQ 753
QY 666 ---GARYRAG-----PGDRRSARVARIQEFRR-RKEREMEELDTLNIRRLPVKMYKG 715
Db 754 PGKSRQKAEQKPPAGKHSSEKSSSSSLAKKRKGEARDCDNKKIRLEKEIKSQ 813
QY 716 HRNSRTMIKEAN 727
Db 814 SSSSSSSSHKSS 825

RESULT 22
US-08-320-559-26
Sequence 26, Application US/08320559
Patent No. 5633135
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaani, Eli

Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LDPKDDTARELKTPSAEERREELRQPPVKRLRGDWSDTGP-----RARPESERERD 332
Db 240 LVAPAPPSQTFPPPSLPKSVAMQOKPTAYVRPM-DGQDQAPSESPKLPDPEDYRQQT 298
QY 333 GEQS-----PNVSLMORMS-----DMLSRW-----SDS 352
Db 299 FEKTDLVKPAKAKTLKMPFSQSVQEQYTSNEVHCVEILKEMTHSWPPPLTAHTPTSTAE 358
QY 353 -----FEASEVAQSNRGRGRPRGGTQSQDISLTPVPSPPDLEVSETAMEVD 402
Db 359 PSKFPPTKDSQHVSVYTNQKQYDTSKTHSNQOCTSSM-----LEDDQLQSDSE--D 411
QY 403 TPAEQFLQ-PSTSTMSAQAHSTSPSPHSTPLSSPDSEQRQSVASGHHTHQSDN 461
Db 412 SDEQPEKPPSSAPPSAPQSLPEPVAHNS-----SSAESTSDSDS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSLHYSTEGTTSTIKLNFTEW-----SSIASSSRGIGS----- 511
Db 461 ESESSSDSENEPLETPAPEPEPTTNKWL-----DNMLTKVSPAPAPPEGPRSTPEPR 517
QY 512 HCKSEQGESFVQ--SSVQPEGDSQETKAPESSE-----DVTKYQEGVSAENPVENHIN 565
Db 518 HPESKSSDSATSQHSESKDPPKSSKAPRAPPEAPHGKRSCQKSPAQOEPQRO-- 575
QY 566 ITQSDKFTAKPL--DSNNGERDNLDRSCGV-----PEESASS 602
Db 576 -TVGTPKPKPKVKASARAGRTSLQGEREGLLPYGRDQTSKDKPKVKTKGRPRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
Db 635 EPKPAVPPSEKKHKSLSLAPSKALSGPEAKDNVEDRTPHEFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLP----- 665
Db 694 SGRTSGRQAVVQEDSRKDRPLPLDRDKLLSPLRDTPPQSLMWKITLDLLSRIPQ 753
QY 666 ---GARYRAG-----PGDRRSARVARIQEFRR-RKEREMEELDTLNIRRLPVKMYKG 715
Db 754 PGKSRQKAEQKPPAGKHSSEKSSSSSLAKKRKGEARDCDNKKIRLEKEIKSQ 813
QY 716 HRNSRTMIKEAN 727
Db 814 SSSSSSSSHKSS 825

RESULT 21
PCT-US94-04496-28
Sequence 28, Application PCT/US9404496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaani, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: For Detection and Treatment of Acute Leukemias
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04496

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

;; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
;; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
;; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
;; TITLE OF INVENTION: All-1 Region
;; NUMBER OF SEQUENCES: 44
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135r1s
;; STREET: One Liberty Place - 46th Floor
;; CITY: Philadelphia
;; STATE: PA
;; COUNTRY: USA
;; ZIP: 19103
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
;; COMPUTER: IBM PS/2
;; OPERATING SYSTEM: PC-DOS
;; SOFTWARE: WORDPERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/320,559
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/062,443
;; FILING DATE: 14 MAY 1993
;; APPLICATION NUMBER: US/07/971,094
;; FILING DATE: 30-OCT-92
;; APPLICATION NUMBER: US/07/888,830
;; FILING DATE: 27-MAY-92
;; APPLICATION NUMBER: US/07/805,093
;; FILING DATE: 11-DEC-91
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DeLuca, Mark
;; REGISTRATION NUMBER: 33,229
;; REFERENCE/DOCKET NUMBER: TJU-0855
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 568-3100
;; TELEFAX: (215) 568-3439
;; INFORMATION FOR SEQ ID NO: 26:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1210 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-320-559-26
Query Match 3.7%; Score 165.5; DB 1; Length 1210;
Best Local Similarity 20.4%; Pred. No. 0.0003;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;
QY 279 LFDPKDDTARELTPSAERREELRQPPVKRLRLRGDWSDTGP-----RARPESERERD 332
DB 263 LVAPAPQPSQTFPPPSLPSKSVAMQPTAYVRPM-DGQQAQPSSELPKPLPDYRQQT 321
QY 333 GEQS-----PVSVMQKMS-----DMLSRW----- 352
DB 322 FEKTDLKVPKAKLTKLMPQSVEQTYSNVHCVEILLKEMTHSWPPLTAHTPSTAE 381
QY 353 -----TEEASEVAQSNRGRSRPRGGTSGQSDSTLPTVPSPDPLEVSTAMEVD 402
DB 382 PSKFPPTKDSQHVSVTQNKQYDTSKTHSNQOQTSSM-----LEDDLQSDSE---D 434
QY 403 TPAEQFLQ-PSTSSMTSAQAHTSPTSPSTPLLSFSDSEKQSVASGHHTHQSDN 461
DB 435 SDEQTFEPKPPSSAPPSAPQLPEFVASAHS-----SSAESTSDSS-----SDS 483
QY 462 NNEKLSPKGTGTPVLSHYEGTGTSTIKLNFTEW-----SSTASSRGTGS----- 511
DB 484 ESESSSDSENEPLETTPAPEPTTNTKWL---DNWLTQVSPAAPPEGPRSTPEPRR 540
QY 512 HCKSEGESESVFPQ---SSVQPPGEDSETKAPDESSE-----DVTKYQEGVSAENPVENHIN 565

DB 541 HPESKSSDSGATSQEHSESKDPPKSSSKAPRAPPEHPGKRSCQKSPAQOEPQOQ--- 598
QY 566 ITQSDKFTAKPL--DNSGERNDLNDRCGV-----PEESASS 602
DB 599 -TVGTQPKPKPVKASARAGRTSLOGEREPGLLPYSGRDOTSKDKPKVKTGPRRAASN 657
QY 603 E-KAKEPETSDOTSTES---ATNENNTNPEFOQ-----TEATGPSAHEE 643
DB 658 EPKPAVPPSSEKKKSKSLPAPSKALSGPEPAKDNVEDRTPEHFALVPLTESQGP-PHSG 716
QY 644 TSTRDSALQDT-----DSDDDDPVLIP----- 665
DB 717 SGRSTSGCRQAVVQVEDSRKDRPLPLRDLKLLSPLRDTPPQSLMVKITLDDLRLSRPQP 776
QY 666 ---GARYBAG-----PGDRRSARAVARQOEFPR-RKERKEMEELDTNIRPLVKNVYKG 715
DB 777 PKGSRQKKAEDKQPPAGKHKHSEKSSSSSLAKKRGAEARDCDNKKIRLEKEIKSQ 836
QY 716 HRNRTMIKEAN 727
DB 837 SSSSSSSSHKES 848
RESULT 23
US-08-545-860D-26
; Sequence 26, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Crocco, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/805,093
: FILING DATE: 11-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Deluca Esq., Mark
: REGISTRATION NUMBER: 33,229
: REFERENCE/DOCKET NUMBER: TJU-1362
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 568-3100
: TELEFAX: (215) 568-3439
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1210 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: IS-08-545-860D-26

```

Query Match 3.7%; Score 165.5; DB 3; Length 1210;
Best Local Similarity 20.4%; Pred. NO. 0.0003;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

```

279 LFDPKDDTARELKTSAEERREELRQPPVKRLRLRGDSDTGP-----RARPESERERD 332
263 LVAPAAPPSQTFPPPSLPSKSVAMQCKTAYVRPM-DGQDQAPSESPELKPDPEDYRQOT 321、
333 GEOS-----PNVSLMORMS-----DMLSRW-----352
322 FEKTDLKVPKAKLTKLAMPQSOSVEQTVSNVHCVTEELKEMTHSWPPLTAHTPSTAE 381
353 -----FEASEVAQSNRGRGRSPRGTSQSDISTLPTVSPSPDLVESETAMEVD 402
382 PSKFPPTKDSQHVSSVTVQKQYDTSSKTHNSOQGTSM-----LEDDQLQSDSE---D 434
403 TPABEQFLO-PSTSTSTMSAQAAHSTSPTESHPSTPLLSPPDSEQRQSVSEAGHHTHQSON 461
435 SDSEQTEPKPPSSNAPSPAQSLPEPVSAHS-----SSAESESTSDSS-----SDS 483
462 NNEKLSKPGTGEVLSLHSTEGTTTSTIKLNFDTBW-----SSIASSSRGIGS-----511
484 ESESSSDSENEPLETPAPEPPTTKNWQL---DNWLTKVSPQAAPEGRSTPEPPR 540
512 HCKSEGOEGSFVPQ---SSVOPPEGDSCTKAPESSE-----DVTKYOGVSAENPVENHJN 565
541 HPESKGSDDSATSOEHSESKDPPKSSKAPRAPPHPGRKSCQKSPAQEPFPQO---598
566 ITQSDKFTAKPL---DSNSCERNDLNDRGCV-----PESASS 602
599 -TVGTQPKPKPVKASRAGRTSLQGERPGLPYGSRDQTSKDKPKVKTKGRPRAASN 657
603 E-RAKEPETSQDSTGS-----ATNENNTNPEPQO-----TEATGSAHEE 643
658 EPKAPVPPSEKKKKHSSLPAPSKALSGPEAKDNVEQTPHFALVPLTESQGP-PHSG 716
644 TSTRDSALQDT---DSDDDDPVLP-----765
717 SGRSTSGCRQAVVQEDSKRDLPLPDRTKLLSLPLRDTPPPSLMVWKITLDSLRIQP 776
666 ---GARYRAG-----PGDRRSAGVARIQEFRR-RKERKEMEELDTLNIRPLVKMYVG 715
777 PKGSGRQKAEQKPPACKKHSSEKSSRSSDSSSLAKRKGAEARCDCKKTRLEKEIKSQ 836
716 HNSRTMIKEAN 727
837 SSSSSSSSHKSS 848

```

```

RESULT 24
PCT-US94-04496-26
; Sequence 26, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli

```

TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
 TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
 TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
 NUMBER OF SEQUENCES: 86
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
 ADDRESSER: Norris
 STREET: One Liberty Place, 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/04496
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Deluca Esq., Mark
 REGISTRATION NUMBER: 33,229
 REFERENCE/DOCKET NUMBER: TJU-1242
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 568-3100
 TELEFAX: (215) 568-3439
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1210 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-04496-26

Query Match	3.7%;	Score 165.5;	DB 5;	Length 1210;
Best Local Similarity	20.4%;	Pred. No. 0.0003;		
Matches 125;	Conservative	80;	Mismatches 218;	Indels 189;
Gaps	26;			
QY	279	LFDPKDDTARELKT	PSAAERRELRQPPV	KRLRGDWSGTG-----RARPESERERD 332
Db	263	LVAPQPPSQTFPP	SLPSKSVAMQOKTAY	VRPM-DGQQAQSESPELKLPELYROOT 321
QY	333	GBOQS-----	PNVSLMQRMS-----	DMLSRW-----352
Db	322	FEKTDLKVPKAK	LKTKLMPQSQVEGY	YNEVHCVEILKEMTHSWPLPTAIHTPSTAE 381
QY	353	-----	FEAESEVAGNSRGR	SRGRGTSQSDISTLPVPPSPDLEVSTAMEVD 402
Db	382	PSKFPPTKDS	OHVSVTONQOYD	TFSSKTHSNSQOGTSSM---LBDDIQLSDSE--D 434
QY	403	TPAEQFLO-PT	SSSTMSQAQAHST	SPSTSPHSTPLSSPDSEQRQSVAEASGHHTHQSODN 461
Db	435	SDSEQTEPKP	SSSAPPAPQQLPE	FAVAHS---SSAESESTSDSS-----SDS 483
QY	462	NNEKLSKPGTGE	PVLSLHYSTEGT	TTTSLKLNFTDEW---SSIASRRGIGS-----511
Db	484	ESESSSDSENE	PLETAPAEPPPT	TNKWL---DNWLTKVSOAPAPPEGRSTPPRR 540
QY	512	HCKSEGOEESVP	Q---SSVQPEGDSE	TKAPEESSE---DVTKTQEGVSAENPVENHIN 565
Db	541	HPESKGSDS	SATSQEHSESKD	PPPKSSSKARAPPEAPHGPKRSCKQSPAQEPPOQ--598
QY	566	ITOSDKFTAK	PL---DSNSGERND	LMLDRSCGV-----PEESASS 602
Db	599	-TVGTKQPKP	VPKASARAGST	SLQGERPGLLYGSRDQTSKDKPKVKTKGRPRAAASN 657
QY	603	E-RAKEPETSD	QSTES---ATTNNNT	NPEPQFQ-----TEATGSAHEE 643
Db	658	EKPAPVPPS	SEKKKHKSSLP	APSKALSGPEPAKDNVEDRTEHFAVLPITBSQGP-PHSG 716
QY	644	TSTRSALQDT	---DSDDDDPV	LIP-----665

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

112 VSCGSGVIEYTNVEODATNROCQFTCHYGTYYEIMTVNDPPTFLSCGDCQTVRW---168
402 SENKDAKAVSLLNNDAEYKYCD--CKHTTTL-VKSVLN-----GNDNTIKEKE 450
169 ---PDTRIKTSCTKE-----DKDDILLNCRRAATSAICPPPIPYLAVGCDSSV 216
451 HIDLDDFSKFGCDKNVDNTNTKVMECKNPYIL-----STKDVCPVPPRQELCIGNID---502
217 RIYDRML-----GTRATGN 231
503 RIYDKNLLMIKEHILATAIYESRILKRYKNKDDKEVCKIINKTFADIRDIIGDTYND 562
232 YAGRTTGMV---ARFIPSHLNKSCR------TSICYSEDG 265
563 LSNRKLVGKINTSKYV--HRNKKNDKLFDRDEWVKVKKOYVNVISWVFKDKTVCKEDDI 620
266 QEILVSYSSDYIYLPDKDDTARELKTSAFERREELPQPVK-----RLRLGDM-SD 318
621 ENI-----POFRWFSEWDDYQDKTKMIETLVKECKEPCEDDNCCKSCSYKEWISK 675
319 TGPRAPESERERDEQSPNVSLMQMSDM-----LSRWFEASEVAQSNRGRSRPRG 373
676 KKEEYNKQAKQOEYQKGNMYSEFKSIKPEVYLKRYSEKSNLNEFEDEFKE-730
374 GTSQSDISTLTPVSPDLEVSETAMEVDTPAEQFLOPSTSTMSAQAHSTSP--TSP 431
731 --LHSDYKNKCTM--CP--EVKDVPLSIIRNNEQTSQEAPEENTELAHRTETPSISEG 784
432 HSTPLLSSPDSEQORQSVASGHHTHQSDNNNEKLSPKGTGEPVLSLHYSTEGITTSI 491
785 -----KGNEQKERDDDSLSKISVSPENSRP-----ETDAKDTNML 819
492 -----KLNEDWSTASSSRGIGSHCKSEGOESFVPOSSVQ 529
820 LKLKGDVDISMPKAVIGSSPNDNINTEQGDNIS-----GVNSKPLSD-----DVR 865
530 PPEGDSFTKAPESSESDVTYKQVSAENPVENHINTQDKFTAKPLDSNG-----ER 584
866 PDKKELEDQNSDESEIVVNH--ISKSPINNGDDSGSGSATVSSSSNTGLSDDDD 922
585 NDMLDRSCGVPEESASSEKAKETSDOTSTES-----ATNENNTPPEQFQTEATG 637
923 NGDTFVRT-----QDTANTVEDIRKENADKDEKGADEERHSTSESLSPEKMLTDNEG 978
638 PSA--HEE---TSTRDSA-----LODT-----DD-----656
979 GNSLNHEEVKEHTSDNDVQQSGGIVNMNVEKLDKTLNPSSSLDEGKAHEELSEPNLS 1038
657 SDDDPVLIPGARYRAGPGRBSAVARIOEFFRR--RKERKEMEELDTLNRRLPVKMYK 714
1039 SDQDMSNTP-----GPLDNTS-----EETTERISNNEYKVNEREDERTLTKEYEDIVLK 1087
715 GHNSR-----TMKEANFWANFVMSGDCGHIFWDRHTAHLMLLEAD 760
1088 SHWNRESDDGELYDENSDLSLVNDESE--DAAKMKGNDTSEM---SHNSSQH---IESD 1139
761 -----NHVYVNCLOPHFPDPLASSGIDYDI-----KWSPLESEFIN--798
1140 QOKNDMKTYGDLGTTTHVQNEIS-----VPVTGEIDELKRESKESKI-HKAEERLSHTD 1192
799 -RKLADDEVITRNELMLEETRN 818
1193 IHKINPEDRNSNTHLKDNRN 1213

RESULT 26
US-08-487-826B-4
Sequence 4, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.

112 VSCGSGVIEYTNVEODATNROCQFTCHYGTYYEIMTVNDPPTFLSCGDCQTVRW---168
402 SENKDAKAVSLLNNDAEYKYCD--CKHTTTL-VKSVLN-----GNDNTIKEKE 450
169 ---PDTRIKTSCTKE-----DKDDILLNCRRAATSAICPPPIPYLAVGCDSSV 216
451 HIDLDDFSKFGCDKNVDNTNTKVMECKNPYIL-----STKDVCPVPPRQELCIGNID---502
217 RIYDRML-----GTRATGN 231
503 RIYDKNLLMIKEHILATAIYESRILKRYKNKDDKEVCKIINKTFADIRDIIGDTYND 562
232 YAGRTTGMV---ARFIPSHLNKSCR------TSICYSEDG 265
563 LSNRKLVGKINTSKYV--HRNKKNDKLFDRDEWVKVKKOYVNVISWVFKDKTVCKEDDI 620
266 QEILVSYSSDYIYLPDKDDTARELKTSAFERREELPQPVK-----RLRLGDM-SD 318
621 ENI-----POFRWFSEWDDYQDKTKMIETLVKECKEPCEDDNCCKSCSYKEWISK 675
319 TGPRAPESERERDEQSPNVSLMQMSDM-----LSRWFEASEVAQSNRGRSRPRG 373
676 KKEEYNKQAKQOEYQKGNMYSEFKSIKPEVYLKRYSEKSNLNEFEDEFKE-730
374 GTSQSDISTLTPVSPDLEVSETAMEVDTPAEQFLOPSTSTMSAQAHSTSP--TSP 431
731 --LHSDYKNKCTM--CP--EVKDVPLSIIRNNEQTSQEAPEENTELAHRTETPSISEG 784
432 HSTPLLSSPDSEQORQSVASGHHTHQSDNNNEKLSPKGTGEPVLSLHYSTEGITTSI 491
785 -----KGNEQKERDDDSLSKISVSPENSRP-----ETDAKDTNML 819
492 -----KLNEDWSTASSSRGIGSHCKSEGOESFVPOSSVQ 529
820 LKLKGDVDISMPKAVIGSSPNDNINTEQGDNIS-----GVNSKPLSD-----DVR 865
530 PPEGDSFTKAPESSESDVTYKQVSAENPVENHINTQDKFTAKPLDSNG-----ER 584
866 PDKKELEDQNSDESEIVVNH--ISKSPINNGDDSGSGSATVSSSSNTGLSDDDD 922
585 NDMLDRSCGVPEESASSEKAKETSDOTSTES-----ATNENNTPPEQFQTEATG 637
923 NGDTFVRT-----QDTANTVEDIRKENADKDEKGADEERHSTSESLSPEKMLTDNEG 978
638 PSA--HEE---TSTRDSA-----LODT-----DD-----656
979 GNSLNHEEVKEHTSDNDVQQSGGIVNMNVEKLDKTLNPSSSLDEGKAHEELSEPNLS 1038
657 SDDDPVLIPGARYRAGPGRBSAVARIOEFFRR--RKERKEMEELDTLNRRLPVKMYK 714
1039 SDQDMSNTP-----GPLDNTS-----EETTERISNNEYKVNEREDERTLTKEYEDIVLK 1087
715 GHNSR-----TMKEANFWANFVMSGDCGHIFWDRHTAHLMLLEAD 760
1088 SHWNRESDDGELYDENSDLSLVNDESE--DAAKMKGNDTSEM---SHNSSQH---IESD 1139
761 -----NHVYVNCLOPHFPDPLASSGIDYDI-----KWSPLESEFIN--798
1140 QOKNDMKTYGDLGTTTHVQNEIS-----VPVTGEIDELKRESKESKI-HKAEERLSHTD 1192
799 -RKLADDEVITRNELMLEETRN 818
1193 IHKINPEDRNSNTHLKDNRN 1213

Query Match 3.7%; Score 165.5; DB 2; Length 1435;
Best Local Similarity 18.1%; Pred. No. 0.00039;
Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;
36 EFIOBKLEATLVHDGCVNTIC-----WNDTGEYILSGSDDTKLVLISN 79
290 EPEKLI-WPAMLSKHNINN-CNNIPQELQITQWIKWH--GEFLLERNRSLK-----341
80 PYSRKVLTTIRSG-----HRANFSKAF-----LPCTNDK-----QI 111
342 PKSKCKNNTLYACEKECIDPCMKYRDWLIIRSKFEWHTLSKEYETQKVPKENAENYLIKI 401

US-08-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERITHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

US-08-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERITHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobb Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelsen, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NI121.001CP1
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-08-487-826B-4

Query Match 3.7%; Score 165.5; DB 2; Length 1435;
 Best Local Similarity 18.1%; Pred No. 0.00039;
 Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;

```

QY 36 EFQRLKLEATLVHDCGVNTIC-----WNDTGEYILSGSDTKLVISN 79
Db 290 EFREKL-WEAMLSHKKNNINN-CRNIPOEELQITQWIKWH-GEFLLERNRSLK----- 341
QY 80 PYSRKVLTTIRSG-----HRANISAKF-----LPCNDK-----QI 111
Db 342 PKSKCKNNTLYEACEKIDPCMKYRDWITRSKFEWHTLSKEYETQKVPKNAENYLIKI 401
QY 112 VSCSGDVITYTVNQAETNRQCFTHYGTVEIMTVPNDPYFLSCGEDGTWRV--- 168
Db 402 SENKDAKVSLLNNGDAEYSKYCD-CRHTTTL-VKSVLN-----GNDNTIIEKRE 450
QY 169 ---FDRIRKTSCKE-----DKDDILNCRRAATVAICPPPIYLAAGCSDSSV 216
Db 451 HIDLDFSGCDKNSVDNTKVCCKNPYL-----STKDVCVPPRRQELCLGNID--- 502
QY 217 RYDRML-----GTRATGN 231
Db 503 RYDKNLLMIKEHILAIAIYESRLKRYKKNDDKEVCKIINKTFADIRLIGGTDYWD 562
QY 232 YAGRTGTGMV---ARFIPSHANKSRV-----TSLCYSEDG 265
Db 563 LSNRLKLVKINTNSYV---HRNKKNDKLFDEWVKVKKDWNVLSWPKDKTVCCKEDI 620
QY 266 QELVSYSSDIYLPDKDDTARELKTSAERRELQPPVK-----RLRLGDM-SD 318
Db 621 ENI-----PQFRFWSEWDDCYQDKTKMIFLKECKECPEDDNCCKSKCNYSKEWISK 675
  
```

```

QY 319 TGPRAPESEERDGEQSPNVSLMQRMDSM-----LSRWFEASEVAQSNRGRSRPRG 373
Db 676 KKEENKQAKQYQYQKGNKYMYSEFKSIKPEVYLKYSKCSNLNFEDEFKEE----- 730
QY 374 GTSQSDISTLTPVSSPDLSEVETAMEYDTPAEQFLQPSSTSTMSAQAHSTSSP--TESP 431
Db 731 --LHSDYKNKCTM--CP--EVKDVPISTIRNNEQTSQEAPEENTEIAHRTETPISSEGP 784
QY 432 HSTPLSSPDSEQRQSVASGHHTHQSDNNNEKLSPPKGTGEPVLSLHYSTEGTTTSTI 491
Db 785 -----KNEQKERDDDSLKISVSPENSRP-----ETDAKDTNLSL 819
QY 492 -----KLNFTDEMSSIASSSRGISGSHCKSEGOESFVPOSSVQ 529
Db 820 LKLGVDVDSMPKAVIGSSPNNDINVTGDNIS-----GVNSKPLSD-----DVR 865
QY 530 PPEGDSKETAPESSEEDVTYKQGVSAENPVENHINITQSKFTAKPLDSNSG-----ER 584
Db 866 PDKKELEDQNSDESEETVNVH-----ISKPSINNGDDSGSGSATVSSSSNTGLSIDD 922
QY 585 NDNLDRSCGVPEESASSEKAKPEPETSOTSTES-----ATNENNTNPEPQOTTEATG 637
Db 923 NGDTFVRT-----QDTANTEDVIRKENADKDEKGADEERHSTSESLSPEEKMLTDNEG 978
QY 638 PSA--HEE---TSTRDSA-----LQDT-----DD----- 656
Db 979 GNSLNHEEVKEHTSNSDNVQSGGIYVMNVYKELKDTLENPSSSLDGGKAHELSERNLS 1038
QY 657 SDDPVLIPGARYRAGFGDRRSAVARIQEFFR--RKERKEMEELDTLNTRRPLVMKMYK 714
Db 1039 SQDMSNTP-----GPLDNTS-----EETTERISNNEYKVNREDETLTKEYEDIVLK 1087
QY 715 GHNSR-----TMKEANFNGANFVMSGDCGHIFIWDRHTAEHLMLLEAD 760
Db 1088 SHNRRESDDGELYDENSLSLTVNDESE--DAEAKMKGNDTSEM---SHNSQH---IESD 1139
QY 761 -----NHRVNCLOPHDPDPILAGSGIDYDI-----KIMSPLESRIFN-- 798
Db 1140 QOKNDMKTGVDLGTTHVQNEIS-----VPVTGEIDEKLRESKESKI-HKAEERLSHTD 1192
QY 799 -RKLADSVITRNELMLEETRN 818
Db 1193 IHKINPEDRNSNTLHLKDRI 1213
  
```

RESULT 27

Sequence 4, Application US/09210288

Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288

585 NDNLDRSCGVPESASSEKAKPEPSTSDSTES-----ATNENTNPFPQFQTEATG 637
 923 NGDTFVKT-----QNTANTEDVIRKENADKDEDEKGADEERHSTSESLSPEKMLTDNEG 978
 638 PSA--HEE-----TSTRDSA-----LQDT-----DD-----656
 979 GNSLNHEEVKEHTSNSDNVQSGGIVNMNVEKELKDTLENPSSSLDEGKAHELSPEPNS 1038
 657 SDDPVLIPGARYAGPDRRSARVIOEFRR--RKERKEMEELDTLNIRRLPLVMVYK 714
 1039 SDQMSNTP-----GPDNTS-----EETTERISNNEYKVNEREDERTLTKEYEDIVLK 1087
 715 GHNSR-----TMKEANFWGANFVMSGDCGHIFIDRHTAHLMLLEAD 760
 1088 SHMNRSDGELYDENSJLSTVNDESE--DAEAKMKND7SEM---SHNSQH---TESD 1139
 761 -----NHVNCLOPHPPDFPILASSGIDYDI-----KIWSPLEESRIFN-- 798
 1140 QOKNDMKTVDLGTTHVQNEIS-----VPVTGEIDKLRRESKESKI-HKAEERLSHTD 1192
 799 -RKLADDEVITRNLMLLETRN 818
 1193 IHKINPEDRNSNTLHLKDIRN 1213

RESULT 28
 US-08-961-083-94
 ; Sequence 94, Application us/08961083
 ; Patent No. 6159469
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 94:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 278 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-961-083-94

Query Match 3.6%; Score 164; DB 4; Length 278;
 Best Local Similarity 22.2%; Pred. No. 4.2e-05;
 Matches 66; Conservative 50; Mismatches 117; Indels 64; Gaps 10;

387 PSSPDLEVSETA-----MEVDTPAEQFLOPSTSTMSQAHSSTSPTE 429

FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-09-210-288-4

Query Match 3.7%; Score 165.5; DB 4; Length 1435;
 Best Local Similarity 18.1%; Pred. No. 0.00039;
 Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;

36 EFIOQLKLEATLVNVDGCVNTIC-----WNTGEXYLSGSDDTKLVISN 79
 290 EFREKL-WEAMLSBKNNINN-CNKIPQEBELQITQWIKWH--GEFLERDNRSKL-----341
 80 PYSRKVLTTIRSG-----HRANISAKF-----LPCNDK-----QI 111
 342 PKSKCKNNLTVEACEKECIDPCMYRWILRSKEFWHTLSKEYETQKVPKRENAENYLIK 401
 112 VSCSGDGVITYNVDOAETNRQCOFTCHGYTTEIMTVPNPDTYFLSCGSDGVTVW---168
 402 SENKNDAKVSLLNCAEAYSKYCD--CKHTTTL-VKSVLN-----GNDNTIKERE 450
 169 ---FTRIKTSCTKE-----DCKDDILNCRRAATSVACPPPIPYILAVGCSDDSV 216
 451 HDLDDFSKFGDKNSVDTNTKWECKNPYL-----STKDCVPPRRQELCLGNID---502
 217 RIYDRML-----HRANISAKF-----LPCNDK-----QI 111
 503 RIYDKNLLMIKEHILAIAIESRLKRYKKNDDKEVCCKIINKTFADIRIIGDTIWDND 562
 232 YAGRTGTGMV---ARFIPSHLNKSCRV-----TSLCYSDG 265
 563 LSNRLVIGKINTNSKIV--HRNKKNDKLFDEWKKVKKOVNWNVISWVFKDKTVCKEDDI 620
 266 QETLVSYSDYIYLFDPKDDTARELKTSAERREELQPPVK-----RLRLGDM--SD 318
 621 ENI-----PQFFRWFWSEWGDYCDQKTKMIETLKYECKEPCEDDCKNSCKNSYKEWISK 675
 319 TGPRAPESERERDQSQNSVLSMQMSDM-----LSRWFEEASEVAOSNRGRGRSRPRG 373
 676 KKEYNKQAKQYQYQKGNMYKMYSEFKIKEVYLKYSKCSNLNFEDEFKEE-----730
 374 GTSQSDISTLPTVPSPLDEVSSTAMEVTPAPQFLOPSTSTMSQAHSSTSP--TESP 431
 731 --LHSDYKNNCTM--CP--EVKDVPIISIRNNEQTSQEAPEENTIAHRTETPISISEGP 784
 432 HSTPILLSPPSEQRQVSQVSEAGHHTHQSNNNEKLPKPGTGPVLSLHYSTEGTTSTI 491
 785 -----KGNQKERRDDUSLSKISVSPNSRP-----ETDAKDTSNL 819
 492 -----KLNFTDEWSSITASSRGGISHCKSEGBQESVPOSSVQ 529
 820 LKLGVDYDIMPXKAVIGSSPNDNINVTQEDNIS---GVNSKPLSD-----DVR 865
 530 PPEGSETKAPESSESDVTKYQGVCSAENPVENHINITQSDKFTAKPLDSNSG-----ER 584
 866 PDKLEQDQNSDESEETVNH---ISKSPSINNGDDSGGSATVSVSSSSNTGLSDDDR 922

```

Db 9 PESPDTVWSDKGEPEQVAPLPEYKGNIEQVKTPEVTEKTEQGPKEEVEV-----PVK 61
QY 430 SPHSTPLSSPDSEORQSVASGHHTHHQSDNNNEKLSPKPGCTGEPVLSLHYSTEGTT 487
Db 62 PTEETPVNNEGTEGTSIQAEANPVQPAEESTTNSKRVSP-----DTSSKN 108
QY 488 TSTIKLNTDEWSSSTASSRGIGSHCKSGEQEESFVQSSVQPPGDSSTKAPESSEDEV 547
Db 109 TGEVSSNPSDSTSGESKNKPHNDSKNENSEKT-VEEVVNPNEGTVBGTSNQETEKPV 167
QY 548 TKYQGVGAENPVENHINITQSDKFTAKPLDSNGERDNLNDRSCGVPPEASSEK --- 604
Db 168 OPAEE-----TQFN--SGKIANENTGEVSKPDSKPPV-EESNOPEKNGT 210
QY 605 AKEPETSQOTSTESATNNTNPEQFQTEATGFSAAHETSTRSALOQTDSDDDP 661
Db 211 ATKPENGNTTSEN-----GOTEPEPSNGNSTEDVSTESNTSNGNEIKQENELDP 263

RESULT 29
US-08-190-802A-63
; Sequence 63, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190.802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUP1, Fig. 46
;
Query Match 3.4%; Score 154; DB 1; Length 713;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 93; Conservative 77; Mismatches 186; Indels 104; Gaps 18;
QY 369 SRPRGGTQSQDSTLPVPSPPDEVTAEVDTPEAQFLQSTSTMSAQHSTSPST 428
Db 159 SRPNLVGSLPTTLPLVSSNAQQQLPQQQLQOQQOQQPPQVSVAPLSNTAINGSPT 218
US-08-190-802A-63

```

```

QY 429 ESPHST-PLSSPDSEORQSVASGHHTHHQSDNNNEKLSPKPGCTGEPVLSLHYSTEGTT 487
Db 219 SKETTILPSVKAPESLTKET-----EPENNTSKIN---DTGSA-----TATTAT 260
QY 488 TSTIKLNTDEWSSSTASSRGIGSHCKSGEQEESFVQSSVQPPGDSSTKAPESSEDEV 547
Db 261 TATETIKPEEDATPASLH-----QDHYLPVYNQRAHNSKPIPPFFLLDSDQS 537
QY 538 ---KAPEESSEDEVTKYQGVGAENPVENHINITQSDKFTAKPLDSNGERDNLN 589
Db 310 VPDALKKQINDYIILNPAIPREIDVELHKSIDHTSVVCCVKF-----SNDGEYLATGC 363
QY 590 DRSCGVPPEASSEKAKEPETSQOTSTESATNNTNPEQFQTEATGFSAAHETSTRS 649
Db 364 NKTTOVYRVSDGSLVARLSDGSAANNHRNSITENTTTSTDNNTWTTTTTTTTTTTAMTS 423
QY 650 ALQ-----DTDDSDDDPVLI-----PGARYRA-GPGDRRSAAVARIOEFRRRK 694
Db 424 AAELAKDVENLTSSSPSDLYIRVCFSPDGKFLATGAEDR---LIRIWDIENRK--- 476
QY 695 EMBELDTLNIIRPLVKMYKGRNSRTMIKEANFWGANF-----VMSGSDCGHIFWDR 748
Db 477 -----IVMILQGH-----EQDIYSLDYFFPSGDKLVSGSDRTVRIWDL 514
QY 749 HTAEHLMLLEADNHVNVNCLQHPDPPIILASSGIDYDIKIW 788
Db 515 RTGCSLTLSTEDGVTTVAVSPGDKYIAAGSLDRAVRVW 554

RESULT 30
US-08-477-346-63
; Sequence 63, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO

```


TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 713 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TUPL1, Fig. 46
US-08-473-089-63

Query Match 3.4%; Score 154; DB 4; Length 713;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 93; Conservative 77; Mismatches 186; Indels 104; Gaps 18;

QY 369 SRPGGTQSQSDISTLTPVSSPDLEVSETAMEVDTPAEQFLQPSSTSTMSAQAHSTSSPT 428
DB 159 SRPNLVGSQSLPTTLVPVSSNAQQQLPQOQLQOQLOQQPPQVSVAPLSNTAINGSPT 218
QY 429 ESPHST-PLLSPPDSEQRQSVASGHHTHHOSDNNNEKLSKPGTGPVLSLHYSTEGTT 487
DB 219 SKETITLPSVKAPESLTKET-----EPENNNTSKIN---DTGSA-----TTATTT 260
QY 488 TST-IKLNFTDEWSSIASRRGIGSHCKSEGOEESFVP-----QSSVQPP---EGDSET 537
DB 261 TATETIKPEEDATPASLH-----QDHYLYPNQRANHSKPIPPFLDLDSQS 309
QY 538 ---KAPEESSEDVTKYQEGVSAENPVE-----NHINITOSDKFTAKPLDSNSGERNDLNL 589
DB 310 VPDALKKQNDYIILNPALPREIDVELHKSLDHTSVCCVKF-----SNDGEYLATGC 363
QY 590 DRSCGVPEESASSEKAKEPTSDQTSATNNTNNTPPEQFOTEGATGPSAHEETSTRDS 649
DB 364 NKTQVYRVSDGSLVARLSDSAAANNHRNSITENNTTSTDNNTMTTTTITTTTAMTS 423
QY 650 ALQ-----DTDDSDDDPVL-----PGARYA-GPGDRRSARVARIQEFRRKERK 694
DB 424 AELAKDVENLNTSSPSDLVRSVCFSPDGKFLATGAEDR---LIRWDIENRK--- 476
QY 695 EMEELDTLNRPLRYKMYKGHRNSRTMIKEANFWGANF-----VMSGDCGHIFWDR 748
DB 477 -----IVMLOGH-----EQDIYSLDYFSPGDKLVSGSGDRTVRIWDL 514
QY 749 HTAEHLMLLEADNHVNCLOPHFPDPFILASSGIDYDIKIW 788
DB 515 RTGQCSLTLSIEDGVTTVAVSPGDKGYIAAGSLDRAVRVW 554

RESULT 32
US-08-487-072A-63
Sequence 63, Application US/08487072A
Patent No. 6423684
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
TITLE OF INVENTION: WD-40 - Derived peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/473,089
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2550-0025.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500

APPLICATION NUMBER: US/08/487,072A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 2550-0025.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 887-1500
 TELEFAX: (202) 887-0763
 INFORMATION FOR SEQ ID NO: 63:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 713 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: TUP1, Fig. 46
 US-08-487-072A-63

Query Match 3.4%; Score 154; DB 4; Length 713;
 Best Local Similarity 20.2%; Pred. No. 0.0011;
 Matches 93; Conservative 77; Mismatches 186; Indels 104; Gaps 18;

```

QY 369 SRPRGGTSSQSDISTLTPTVSPDLEVSAMEVDTFAEQFLQSTSTMSAQAHSTSSPT 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 SRPNLVGSQLPTTLVPSVSSNAQQQLPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 218

QY 429 ESPHST-PLLSPPDSQROSVASGHHHTHQSDNNNEKLSPKPGTGPVLSLHYSTEGIT 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 SKETTLPSVKAPESTLKET-----EPENNTSKIN---DTGSA-----TTATTT 260

QY 488 TST-IKLNFTDEWSSIASRRGIGSHCKSEGESEFVP-----QSSVQPP-----EGDSET 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TATEIKPKEDATPASLH-----QDHYLVYPYNQRANHSKPIPPFLDLDSQS 309

QY 538 ---KAPESEEDVTKYQEGVSAENPVE-----NHINITQSDKFTAKPLDSNGERNDLNL 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 VPDALKKQTDYILYNALPREIDVELHKSLDHTSVVCCVKF-----SNDGEYLATGC 363

QY 590 DRSCGVPEESASSEKAKEPETSDQSTESATNNTNPEPQFQTEATGPSAHEETSTRDS 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 NKTQVYRVSQGLSVARLSDDSAANNHRNSITENNTTTSTDNNTMTTTTITTTTAMTS 423

QY 650 ALQ-----DTDDSDDDPVLII-----PGARYA--GPGDRRSAVARIOEFFFFRRKERK 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 AAELAKDVENLNTSSPSDDLIRSVCFSPDGKFLATGAEDR---LIRWDIENRK----- 476

QY 695 EMEELDTLNIIRPLVKVYKGRNSRTWIKENFWGANF-----VMSGSDCGHIFWDR 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 -----IVMILQGH-----EQDIYSLDYFPPSGDKLVSGSGDRVTRIWDL 514

QY 749 HTAHLMLLEADNHVNCLOPHFPDFPIASSGIDYDIKIW 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 RTGQCSLTLSIEDGVTTVAVSPDGKGYIAAGSLDRAVRVW 554

```

RESULT 33
 US-09-108-857-3
 Sequence 3, Application US/09108857A
 Patent No. 6433137
 GENERAL INFORMATION:
 APPLICANT: JOHNSON, Alexander D.
 APPLICANT: BRAUN, Burkhard
 TITLE OF INVENTION: TUP1 SEQUENCES FROM CANDIDA ALBICANS AND METHODS FOR
 TITLE OF INVENTION: SCREENING AGENTS FOR INHIBITING VIRULENCE IN CANDIDA
 TITLE OF INVENTION: ALBICANS USING TUP1
 FILE REFERENCE: 22002-20006.00
 CURRENT APPLICATION NUMBER: US/09/108,857A
 CURRENT FILING DATE: 1998-07-01
 EARLIER APPLICATION NUMBER: 60/051,552

EARLIER FILING DATE: 1997-07-02
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 3
 LENGTH: 713
 TYPE: PRT
 ORGANISM: Candida albicans
 US-09-108-857-3

Query Match 3.4%; Score 154; DB 4; Length 713;
 Best Local Similarity 20.2%; Pred. No. 0.0011;
 Matches 93; Conservative 77; Mismatches 186; Indels 104; Gaps 18;

```

QY 369 SRPRGGTSSQSDISTLTPTVSPDLEVSAMEVDTFAEQFLQSTSTMSAQAHSTSSPT 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 SRPNLVGSQLPTTLVPSVSSNAQQQLPQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 218

QY 429 ESPHST-PLLSPPDSQROSVASGHHHTHQSDNNNEKLSPKPGTGPVLSLHYSTEGIT 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 SKETTLPSVKAPESTLKET-----EPENNTSKIN---DTGSA-----TTATTT 260

QY 488 TST-IKLNFTDEWSSIASRRGIGSHCKSEGESEFVP-----QSSVQPP-----EGDSET 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TATEIKPKEDATPASLH-----QDHYLVYPYNQRANHSKPIPPFLDLDSQS 309

QY 538 ---KAPESEEDVTKYQEGVSAENPVE-----NHINITQSDKFTAKPLDSNGERNDLNL 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 VPDALKKQTDYILYNALPREIDVELHKSLDHTSVVCCVKF-----SNDGEYLATGC 363

QY 590 DRSCGVPEESASSEKAKEPETSDQSTESATNNTNPEPQFQTEATGPSAHEETSTRDS 649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 NKTQVYRVSQGLSVARLSDDSAANNHRNSITENNTTTSTDNNTMTTTTITTTTAMTS 423

QY 650 ALQ-----DTDDSDDDPVLII-----PGARYA--GPGDRRSAVARIOEFFFFRRKERK 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 AAELAKDVENLNTSSPSDDLIRSVCFSPDGKFLATGAEDR---LIRWDIENRK----- 476

QY 695 EMEELDTLNIIRPLVKVYKGRNSRTWIKENFWGANF-----VMSGSDCGHIFWDR 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 -----IVMILQGH-----EQDIYSLDYFPPSGDKLVSGSGDRVTRIWDL 514

QY 749 HTAHLMLLEADNHVNCLOPHFPDFPIASSGIDYDIKIW 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 RTGQCSLTLSIEDGVTTVAVSPDGKGYIAAGSLDRAVRVW 554

```

RESULT 34
 PCT-US93-03077-1
 Sequence 1, Application PC/TUS9303077
 GENERAL INFORMATION:
 APPLICANT: Board of Regents, The University of Texas System
 APPLICANT: Gaynor, Richard B.
 APPLICANT: Wu, Foon Kin
 TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
 TITLE OF INVENTION: REGULATING GENE EXPRESSION
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Arnold White & Durkee
 STREET: P.O. Box 4433
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/03077
 FILING DATE: 19930331
 CLASSIFICATION:
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/862,025
FILING DATE: April 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: UTFD270PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1540
TELEFAX: 713-749-2679
TELEFAX: 713-749-2679
TELEFAX: 713-749-2679

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1093 amino acids
;   TYPE: AMINO ACID
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: protein
; PCT-US93-03077-1
;
; Query Match 3.4%; Score 154; DB 5; Length 1093;
; Best Local Similarity 20.3%; Pred. No. 0.0021;
; Matches 100; Conservative 81; Mismatches 186; Indels 126; Gaps 20;
;
; QY 269 LVYSQSDYIYLFDFKDDTARELKTPTSAERRELROPPVKRLRLRGDWDSDTGRPRPESE 328
;   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
; Db 89 VYDESNFFSFLSTDVQTQKSP-----VVSFPAK-----SQRPSEE 128
;
; QY 329 RERDGEQSPNYSLMORMSDMLSRFDEPASEVAOSNRGRHSRRPRG-----TSQSDIS 381
;   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
; QY 128 VFSSTVFSFHIGOSPTPTTSSQVSDSLCVSGETLAAQTSSPKTEGHEETVNNKSDMK 188
;   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

382	QY	TLPTVPSSPDLEVSFTAMEVDTPAEQFLOPSTSTMSAQAHSTSSSTESPHSTPLSSPD	441
189	Db	VTQV-----SLKVSFVIVKTTMESINTSTQSLTAETKDIALEPKOHH-----	234
442	QY	SEQRQSVASGHHTHQSDNNNEKLSPPGTGEVLSHYSTGTTTSTIKLNFDEWS	501

```
001 YQ   IASSKRIGRKASGQQDQLLDTLSRST :      |      |      |      |      |      |      |      |      |  
    Db  268 ISSESS-----ASSRGQTTSKSLLHMQTSLFQSACPEYNRLDDFOKLTESCCS 319  
  
    QY   ENPVENHNITQSDKFTAKPLDSNGERNDMLDRSCG-----VP-EESASSEKAKEP 609  
    Db  320 SDAFE-----RIDSFVSQGLSDRS--VGEINSDDELCKGYALVPIIVNSSTPKRKIVE 371  
  
    QY   TSDOTSTSATENNNTNPQFOTEATGPSA-----HEETSTRDSLALQ 652  
    Db  372 SAEGKSEE-VNETHVIPTEEAEWESSGSRAFPVNCEQPDIILSVSTPINEGQTVLDKVAE 429  
  
    QY   DTDDSDDDLIPICARYAGPGRRSAVARIQBFPRKKEMELTDIIRPLEVKMV 712  
    Db  430 QCFAESQEPALE-----SEKDGVCTVEFLNEKLEKREAOLL-SLSREKALLEEA 478  
  
    QY   713 YKGHRNSTMIKE 725  
    Db  479 FDLNKDEMFRVKE 491
```

RESULT 35
US-08-190-802A-202
; Sequence 202, Application US/08190802A
; Patent No 5519003
; GENERAL INFORMATION:
; APPLICANT: Moichly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Dehninger & Associates

; REFERENCE/DOCKET NUMBER: 2550-0025.20
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 887-1500
 ; TELEFAX: (202) 887-0763
 ; INFORMATION FOR SEQ ID NO: 202:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 50 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37
 ; US-08-477-346-202

Query Match 3.4%; Score 153.5; DB 4; Length 50;
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAEHLMLLEAD-NHVVNCLQHPFDPILASSGIDYDIKIWS 789
 DB 1 GHIFWEKSCQIQVLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

RESULT 37

US-08-473-089-202
 ; Sequence 202, Application US/08473089
 ; Patent No. 6342368

; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; NUMBER OF INVENTION: Thereof
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 2000 Pennsylvania Avenue, NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/473,089
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
 ; NAME: MURASHIGE, KATE H.
 ; REGISTRATION NUMBER: 29,959
 ; REFERENCE/DOCKET NUMBER: 2550-0025.22
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 887-1500
 ; TELEFAX: (202) 887-0763

; INFORMATION FOR SEQ ID NO: 202:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 50 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37

US-08-473-089-202
 Query Match 3.4%; Score 153.5; DB 4; Length 50;
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAEHLMLLEAD-NHVVNCLQHPFDPILASSGIDYDIKIWS 789
 DB 1 GHIFWEKSCQIQVLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

QY 741 GHIFWDRHTAEHLMLLEAD-NHVVNCLQHPFDPILASSGIDYDIKIWS 789
 DB 1 GHIFWEKSCQIQVLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

RESULT 38

US-08-487-072A-202
 ; Sequence 202, Application US/08487072A
 ; Patent No. 6423684

; GENERAL INFORMATION:
 ; APPLICANT: Mochly-Rosen, Daria
 ; APPLICANT: Ron, Dorit
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
 ; NUMBER OF INVENTION: Thereof
 ; NUMBER OF SEQUENCES: 265
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Morrison & Foerster
 ; STREET: 2000 Pennsylvania Avenue, NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20006-1812

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,072A
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:
 ; NAME: MURASHIGE, KATE H.
 ; REGISTRATION NUMBER: 29,959
 ; REFERENCE/DOCKET NUMBER: 2550-0025.20
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 887-1500

; INFORMATION FOR SEQ ID NO: 202:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 50 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37

US-08-487-072A-202

Query Match 3.4%; Score 153.5; DB 4; Length 50;
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAEHLMLLEAD-NHVVNCLQHPFDPILASSGIDYDIKIWS 789
 DB 1 GHIFWEKSCQIQVLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

RESULT 39

US-08-769-309A-5
 ; Sequence 5, Application US/08769309A
 ; Patent No. 5741890

; GENERAL INFORMATION:
 ; APPLICANT: Scott, John D.,
 ; APPLICANT: Nauert, Brian J.,
 ; APPLICANT: Klauck, Theresa M.
 ; TITLE OF INVENTION: Protein Binding Domains of Gravin
 ; NUMBER OF SEQUENCES: 24
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower/233 South Wacker Drive

QY 771 P-----FDPILASGIDYDIKIW-----SPLSESRIFNRKLA 802
 Db 873 QKAATEVSKELSESQVHMMAAAVADGTRAATIERSPSWISASVTEPLEQVEAEAAALT 932
 QY 803 DEVITRNLMELETRNITVTP 823
 Db 933 EEVLER-EVIAEEEPPTVTEP 952

RESULT 40
 US-08-994-570-5
 : Sequence 5, Application US/08994570
 : Patent No. 6090929
 : GENERAL INFORMATION:
 : APPLICANT: Scott, John D.,
 : APPLICANT: Nauert, Brian J.,
 : APPLICANT: Kluwert, Theresa M.
 : TITLE OF INVENTION: Protein Binding Domains of Gravin
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 : STREET: 6300 Sears Tower/233 South Wacker Drive
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: United States of America
 : ZIP: 60606-6402
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/994.570
 : FILING DATE:
 : CLASSIFICATION:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: No. 6090929and, Greta E.
 : REGISTRATION NUMBER: 35,302
 : REFERENCE/DOCKET NUMBER: 27866/33451
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 312-474-6300
 : TELEFAX: 312-474-0448
 : TELEX: 25-3856
 : INFORMATION FOR SEQ ID NO: 5:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1780 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-769-309A-5

Query Match 3.4%; Score 152.5; DB 1; Length 1780;
 Best Local Similarity 20.7%; Pred. No. 0.0058;
 Matches 141; Conservative 249; Mismatches 249; Indels 205; Gaps 33;

QY 281 DPKDDTARELTPS-----AERREELRPPVKRLRGDWDGTGPRAPESEREDGEQS 336
 Db 339 DGKAEVASEKLTAHQAPQPAESAHEP-----RLSAEYE-----KVLPSEEQVSSQG 389
 QY 337 PNVSLMORMSDMLSNWFEEASVQSNRGRSRPGTQSOSDITLTPVSPDPLEVSE 396
 Db 390 PS-----EERPAPLATEVDFDKIEVHOEE-----VVAEVHVSTVEERTEEQKTEVEE 436
 QY 397 TAMEVDTPAEQFL-----QPSTSS-----TMSAQAHSTSSPTSPHSTP---LLSSPDE 443
 Db 437 TAGSV--PAELVGMDAEPQAEPAKELVKETCVSGEDPTQGADLSPDKVLSKPPBG 494
 QY 444 QROSVASGHHTHQSDNNNEKLSPPGCTGEPVLSLHYSTEGTTTSTIKLNFDEWSSIA 503
 Db 495 VVSEVEML-----SSQERMKVQ---GSLPKLF-----TSTGLKLSGKKQ-----532
 QY 504 SSSRGIGSHCKSGQESFVPSQSVOPPE---GDSETKAPESSE-----545

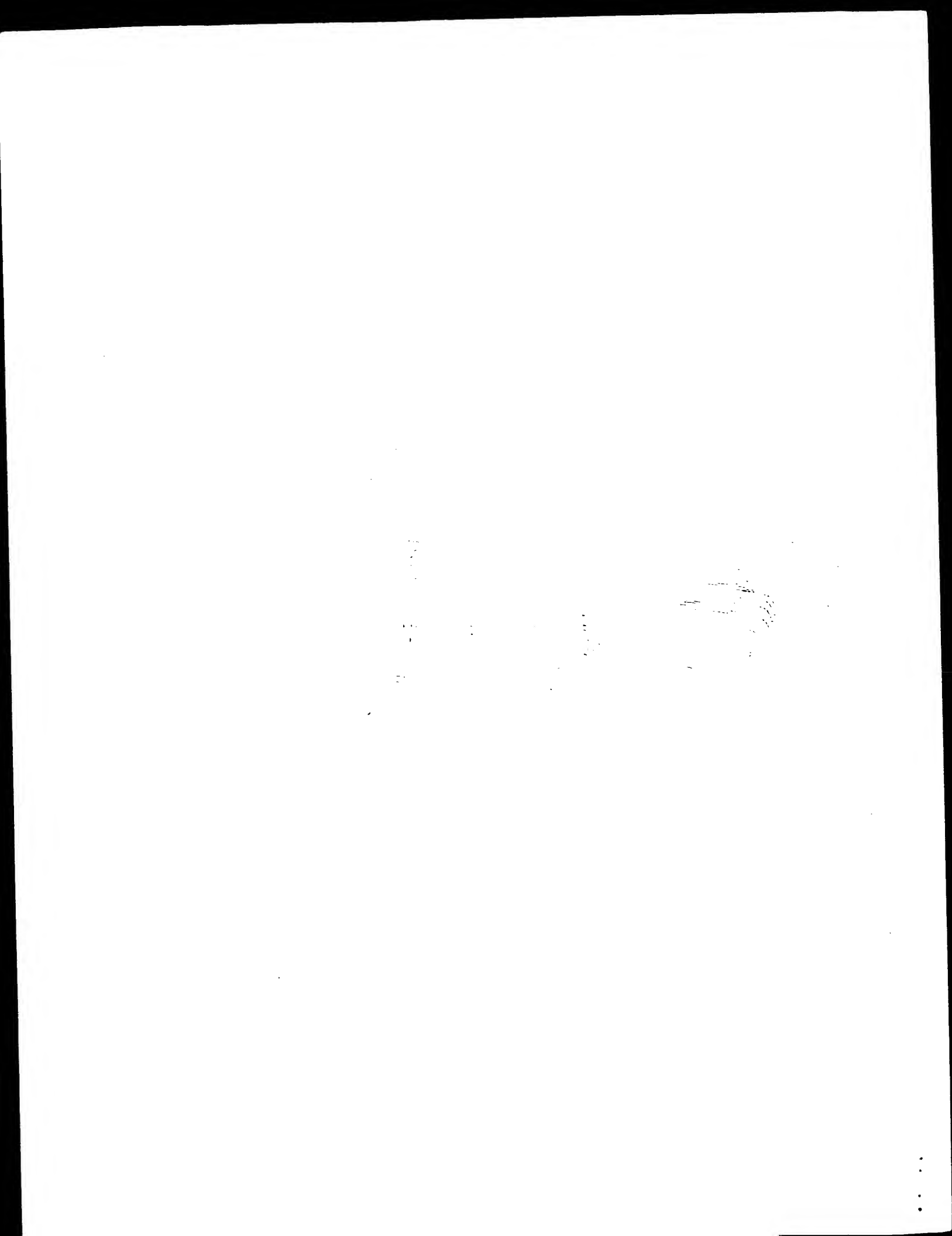
CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/769.309A
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 5741890and, Greta E.
 REGISTRATION NUMBER: 35,302
 REFERENCE/DOCKET NUMBER: 27866/33451
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1780 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-769-309A-5

Query Match 3.4%; Score 152.5; DB 1; Length 1780;
 Best Local Similarity 20.7%; Pred. No. 0.0058;
 Matches 141; Conservative 249; Mismatches 249; Indels 205; Gaps 33;

QY 281 DPKDDTARELTPS-----AERREELRPPVKRLRGDWDGTGPRAPESEREDGEQS 336
 Db 339 DGKAEVASEKLTAHQAPQPAESAHEP-----RLSAEYE-----KVLPSEEQVSSQG 389
 QY 337 PNVSLMORMSDMLSNWFEEASVQSNRGRSRPGTQSOSDITLTPVSPDPLEVSE 396
 Db 390 PS-----EERPAPLATEVDFDKIEVHOEE-----VVAEVHVSTVEERTEEQKTEVEE 436
 QY 397 TAMEVDTPAEQFL-----QPSTSS-----TMSAQAHSTSSPTSPHSTP---LLSSPDE 443
 Db 437 TAGSV--PAELVGMDAEPQAEPAKELVKETCVSGEDPTQGADLSPDKVLSKPPBG 494
 QY 444 QROSVASGHHTHQSDNNNEKLSPPGCTGEPVLSLHYSTEGTTTSTIKLNFDEWSSIA 503
 Db 495 VVSEVEML-----SSQERMKVQ---GSLPKLF-----TSTGLKLSGKKQ-----532
 QY 504 SSSRGIGSHCKSGQESFVPSQSVOPPE---GDSETKAPESSE-----545
 Db 533 KGRGGD---EESGEHTQVPADSPDSQBEQKGESSASSPEEPETCTCLEGLAEVQQDG 589
 QY 546 -----DVTKYQGVSAENPVENHINTQSDKFTAKPLDSNGSRNDLNRSCGVPE 597
 Db 590 EAEGATSDGEKKREGVT---PWASFKWVTPKKVRRP--SESDREDELKVKSATLSS 644
 QY 598 -ESASSE-----KAKEPETSQDST-----ESATNENNTNPEPQFQ 632
 Db 645 TESTASEMOEBMGVSEPEKPEEPKRVDTSVSWEALICVSGSKKRRARRSSD-----698
 QY 633 TEATGPSA---HEETSTRSALQDQDSD-----DPVLIPGARYRAGPGRRSA 679
 Db 699 -EEGFPKAMGDHQA---DEAGDKETGTDGILLGQSEHDPGGGSSPEQAGSPTEGEG 754
 QY 680 VARIOEPR---RRKERKEEELDTNI-----RRP-----LVKMVYKGRH 717
 Db 755 VSTWSEFKRLVTPRKKSKELEKSEDSIAGSGVHSTPDTPEKESWSVSKFIFGR 814
 QY 718 NER-----TWIKANFWANFVMSGSDCGHIFLWDRHTAHLMLLEADNHYVNCLOPH 770
 Db 815 KKRDPGKQAPVEDAGPTGAN--EDDSQVAVVPLSEYDAVEREKMEAAQQAQAGBOPE 872

Db 533 GKRCGGD---EESGEHTQVPADSPDSOEQKQESSASSPEEPEITCLEKGLAEVQODG 589
QY 546 -----DVTKYQGVSAENPVENHINITQDKFTAKPLDSNGERNDLNLDRSCGVPE 597
Db 590 EAEGATSDGKKEGVT---PWASFKKMVTPKRVRP--SESOKEDLDKVKSATLSS 644
QY 598 -ESASSE-----KAKPETSQOTST-----ESATNENNTNPPEQFQ 632
Db 645 TESTASEMOEEMKGVPEPKPEPKRYDTSVWEALICVGSKKRRRRSSD----- 698
QY 633 TEATGPSA---HEETSTRSALQTDSDDD-----DPVLPGARYRAGPGRRSA 679
Db 699 -EGGPKAMGGDHOKA---DEACKKETGTCIILAGSQEHDPGOGSSSPQAGSPTGEG 754
QY 680 VARIOEFFR---RRKERKEMEELDTLNI-----RRP-----LVKMVYKGR 717
Db 755 VSTWSEFKRLVTPRKKSLSKEEKSEDSIAGSGVEHSTPDTEPKESWYSIKKFIPIRR 814
QY 718 NSR-----TMIKEANFWGANFVMSGDCGHIFWDRHTAEHMLLEADNHVNVNCLQPH 770
Db 815 KRPDQKQEQAPVEDAGPTGAN--EDDSVPAVVPLSEYDAVEREKMEAOQAKGAEQPE 872
QY 771 P-----FDPILASSGIDYDIKIW-----SPLAESRIFNRKLA 802
Db 873 OKAATEVSKELSESQVHMAAAVADGTRAATILEERSPSWISASVTEPLEQVEAEALLT 932
QY 803 DEVITRNLMLLETRNTITVP 823
Db 933 EEVLER-EVIAEEEPPTVTEP 952

Search completed: March 18, 2003, 15:36:37
Job time : 59 secs



GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:29:40 ; Search time 15 seconds
(without alignments)
2377.976 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSGGSPHLLWDVRRSLG.....RLGGDRSEGGQENENDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	221.5	4.9	489	1	DMP1_RAT
2	221	4.9	503	1	DMP1_MOUSE
3	198.5	4.4	406	1	SR40_YEAST
4	190.5	4.2	1367	1	ANYH_YEAST
5	189	4.2	1142	1	MGCL_HUMAN
6	188	4.2	1702	1	IGA2_HAETIN
7	186	4.1	1694	1	IGA0_HAETIN
8	184	4.1	1025	1	HIRA_FUGRU
9	177.5	3.9	1658	1	YM67_YEAST
10	175.5	3.9	1849	1	IGA4_HAETIN
11	175	3.9	934	1	DSPP_MOUSE
12	174.5	3.9	787	1	NASP_MOUSE
13	174.5	3.9	2004	1	MOZ_HUMAN
14	173	3.8	2492	1	ATRX_HUMAN
15	172	3.8	510	1	DMP1_BOVIN
16	171.5	3.8	513	1	DMP1_HUMAN
17	171.5	3.8	1189	1	DXT STRO
18	170.5	3.8	1337	1	XJH6_YEAST
19	169.5	3.7	1092	1	NCA2_XENLA
20	169	3.7	1253	1	DSPP_HUMAN
21	168.5	3.7	1389	1	DC13_DROME
22	168	3.7	841	1	IEG3_MCMVS
23	168	3.7	910	1	DNUM_MYCPN
24	167.5	3.7	527	1	ATRX_RAT
25	167.5	3.7	704	1	YM96_YEAST
26	167.5	3.7	1140	1	ATRX_CABEL
27	166.5	3.7	1359	1	PCGV_HUMAN
28	166.5	3.7	3396	1	AF4_HUMAN
29	165.5	3.7	1210	1	EBAL_PLAFC
30	165.5	3.7	1435	1	ANK2_HUMAN
31	165.5	3.7	3924	1	AL1A_CANAL
32	164.5	3.6	1419	1	AL1A_CANAL
33	163	3.6	1466	1	SFA2_YEAST

34	162.5	3.6	699	1	NP14_HUMAN
35	161.5	3.6	671	1	CHS5_YEAST
36	160.5	3.5	1356	1	HET1_PODAN
37	160.5	3.5	4377	1	ANK3_HUMAN
38	160	3.5	480	1	TGN2_HUMAN
39	159.5	3.5	1447	1	BUD4_YEAST
40	159	3.5	1275	1	YAU9_SCHPO
41	159	3.5	1310	1	YB35_SCHPO
42	158	3.5	2774	1	MAPA_RAT
43	157.5	3.5	680	1	NASP_RABIT
44	157.5	3.5	1912	1	APC_CHICK
45	157.5	3.5	2842	1	APC_RAT
46	156.5	3.5	713	1	TUP1_YEAST
47	156	3.4	1015	1	HIRA_MOUSE
48	156	3.4	1070	1	PVDG_PLAKN
49	155.5	3.4	1020	1	NPH_HUMAN
50	155.5	3.4	1217	1	AF4_MOUSE
51	155	3.4	687	1	DSPP_RAT
52	155	3.4	1062	1	SUM1_YEAST
53	154.5	3.4	589	1	HIBN_XENLA
54	154.5	3.4	1095	1	NEB1_RAT
55	154	3.4	1093	1	TMF1_HUMAN
56	154	3.4	2476	1	ATRX_MOUSE
57	153	3.4	725	1	AGAL_YEAST
58	153	3.4	1181	1	NKX1_RAT
59	152.5	3.4	1781	1	AKAC_HUMAN
60	152.5	3.4	1839	1	CYAA_SACKL
61	152	3.4	623	1	HPC2_YEAST
62	151.5	3.3	601	1	YFK5_SCHPO
63	151.5	3.3	1139	1	HMW1_MYCGE
64	151.5	3.3	1403	1	YDF3_SCHPO
65	151.5	3.3	2440	1	NCRI_HUMAN
66	151	3.3	572	1	CAFB_MOUSE
67	151	3.3	1075	1	FSB2_YEAST
68	151	3.3	1306	1	FL05_YEAST
69	150.5	3.3	943	1	CENC_HUMAN
70	150.5	3.3	1233	1	YF16_YEAST
71	150.5	3.3	1972	1	P531_HUMAN
72	150.5	3.3	2453	1	NCRI_MOUSE
73	150	3.3	340	1	GBB1_HOMAM
74	150	3.3	651	1	CORO_YEAST
75	150	3.3	679	1	YIS3_YEAST
76	150	3.3	742	1	CD44_HUMAN
77	149.5	3.3	1322	1	YAG3_YEAST
78	149.5	3.3	1805	1	P3K3_DICDI
79	149.5	3.3	1585	1	NEST_RAT
80	149	3.3	1155	1	PVDB_PLAKN
81	149	3.3	1162	1	TCNA_TRYCR
82	148.5	3.3	562	1	CHS5_CANAL
83	148.5	3.3	633	1	MLH_TETTH
84	148.5	3.3	915	1	NFM_HUMAN
85	148.5	3.3	1229	1	MRSP_STRAU
86	148.5	3.3	1637	1	SLP1_CLOTM
87	148.5	3.3	1664	1	MSPI_PLAYO
88	148.5	3.3	2349	1	TPR_HUMAN
89	148.5	3.3	2805	1	MAPA_HUMAN
90	148.5	3.3	1788	1	YF72_CAEFL
91	148	3.3	2132	1	PGCA_MOUSE
92	148	3.3	427	1	YKTS_YEAST
93	147.5	3.3	856	1	Y249_HUMAN
94	147.5	3.3	2090	1	N214_HUMAN
95	147.5	3.3	2843	1	APC_HUMAN
96	147.5	3.3	3866	1	HRX_MOUSE
97	147.5	3.3	3969	1	HRX_HUMAN
98	147.5	3.3	375	1	PEX7_YEAST
99	147	3.2	377	1	GBB2_TOBAC
100	147	3.2	620	1	CHS5_SCHPO
101	147	3.2	1142	1	PAK1_YEAST
102	147	3.2	1453	1	NKCR_MOUSE
103	147	3.2	1091	1	NCA1_CHICK
104	146.5	3.2	1411	1	Y297_HUMAN
105	146.5	3.2	377	1	GBB_ARATH
106	146	3.2			

Q14978 homo sapien
Q12114 saccharomyc
Q00808 podospora a
Q12955 homo sapien
P47136 saccharomyc
Q10164 schizosacch
Q14340 schizosacch
P34926 rattus norv
P27123 oryctolagus
P87498 gallus gall
P70478 rattus norv
P16649 saccharomyc
Q61666 mus musculu
P50494 plasmodium
P12036 homo sapien
Q88573 mus musculu
Q62598 rattus norv
P46676 saccharomyc
Q61806 xenopus lae
Q35867 rattus norv
P82094 homo sapien
Q61687 mus musculu
P32323 saccharomyc
Q92m6 rattus norv
Q02952 homo sapien
P23466 saccharomyc
Q01448 saccharomyc
P87132 schizosacch
Q49413 mycoplasma
Q10475 schizosacch
Q75376 homo sapien
Q9d0n7 mus musculu
P38894 saccharomyc
P32334 saccharomyc
Q03188 homo sapien
P43597 saccharomyc
Q12888 homo sapien
Q60974 mus musculu
Q05040 homarus ame
Q06440 saccharomyc
Q0563 saccharomyc
P16070 h cd44 anti
P39712 saccharomyc
P54675 dictyosteli
P21263 rattus norv
P50493 plasmodium
P23253 trypanosoma
Q74161 candida alb
Q40631 tetrahymena
Q07197 homo sapien
Q9Y2n3 homo sapien
P80544 staphylococ
Q06852 ciostidium
P13828 plasmodium
P12270 homo sapien
P78559 homo sapien
Q92221 caenorhabdi
Q61282 mus musculu
P36046 saccharomyc
Q92539 homo sapien
P35658 homo sapien
P25034 homo sapien
P55200 mus musculu
Q30164 homo sapien
P39108 saccharomyc
P33398 nicotiana t
Q92357 schizosacch
P38990 saccharomyc
P30415 mus musculu
P13590 gallus gall
Q15040 homo sapien
P49177 arabidopsis

Wed Mar 26 09:38:55 2003

107 146 3.2 1087 1 NPH_MOUSE
108 146 3.2 1537 1 FLOL_YEAST
109 145.5 3.2 361 1 WDS_DROME
110 145.5 3.2 700 1 TRDN_CANFA
111 145.5 3.2 1395 1 SP41_YEAST
112 145.5 3.2 1403 1 CIC_DROME
113 145 3.2 377 1 GBB_SOLTU
114 145 3.2 614 1 TULL_SCHPO
115 145 3.2 1088 1 NCAL_XENLA
116 145 3.2 2845 1 APC_MOUSE
117 144 3.2 482 1 YSR2_CAEEL
118 144 3.2 501 1 YH92_CAEEL
119 144 3.2 1309 1 RAD9_YEAST
120 143.5 3.2 870 1 Y563_HUMAN

ALIGNMENTS

RESULT 1
DMP1_MOUSE STANDARD; PRT; 489 AA.

AC P98193;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
protein-1) (DMP-1) (AGI).

GN DMP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Tooth;
RX MEDLINE=93286101; PubMed=8509401;
RA George A., Sabsay A., Simonian P.A., Veis A.;
RT "Characterization of a novel dentin matrix acidic phosphoprotein.
Implications for induction of biomineralization."
J. Biol. Chem. 268:12624-12630(1993).
RL -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF
EXTRACELLULAR MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN ODONTOBLAST
AND AMELOBLAST.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; L11354; -; NOT ANNOTATED_CDS.
CC Extracellular matrix; Signal.
CC SIGNAL 1 16 POTENTIAL.
CC CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
CC DOMAIN 41 44 POLY-PRO.
CC FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).
CC SEQUENCE 489 AA; 53058 MW; 59F8381479DDA085 CRC64;

Query Match 4.9%; Score 221.5; DB 1; Length 489;
Best Local Similarity 25.9%; Pred. No. 0.00018;
Matches 103; Conservative 56; Mismatches 164; Indels 75; Gaps 14;

QY 281 DPKDDT-ARELUTPAERREELRQPPVKRLRLRGDQSDTGPAPRPSERERDGEQSPNV 339
DB 89 DSGDDTDFGDDNGPGPEER----QWGGPSRLSDSDSDTQSSDSTSQNSAQDTPSD 144

QY 340 SLQMDSMLSRFEAEVAQ---SNRGRSRPRGGTSQSDISTLPTVPSSPDLEVSE 396
DB 145 SKDHSDEADR--PEAGDSTQDSESEYRVGGSEGESHGDCS-----EFDD 191
QY 397 TAMEVDTPAQQLQPTSTSTMSQAQHSSTPT-----ESPSTPLSSSPDEQORQSVAS 451
DB 192 EGMQSDDP-----GSTRSDRGHTRMSAGIRSEESKGDHEPTSTQSDSDSODVEFS 242
QY 452 GHYTHQSDNNNEKLSKPGTGPVLSLHYSTGTSTIKLNTDEWSSIASSSRGIGS 511
DB 243 SRKSFRRS-----RVSEDDRGELADNSRSTQSVSTDFRSK-----E 281
QY 512 HCKSGQSESVPOSSQVPPPEG-----DSETKAPEESEDVTKYQEGVSAENPVENHINI 566
DB 282 ERSSTQEDTAETQSDQSDPEQDPSPSESEAEAGEPQSESSQEGVASESRGNDPNT 341
QY 567 TQSDKFTAKPLDSNGERNDLNRSCGVPPEASASSEKAKEPETS-----QTSSTSATN 621
DB 342 SQ---TGQRDSSESEEDRLNTFSS-----SESQSTEQGDSNESLSLSEESQESAQD 393
QY 622 ENNTNPEP-QFQTEATGPSAHEETSTRDSALQDITDSD 658
DB 394 EDSQSGEGIQSQSASRESRSQESQSDSRSENRDSD 431

RESULT 2

DMP1_MOUSE STANDARD; PRT; 503 AA.

AC O55188;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix
protein-1) (DMP-1) (AGI).
DE DMP1 OR DMP.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Swiss Webster; TISSUE-Molar;
RX MEDLINE=98184255; PubMed=9525343;
RA McDougall M., Gu T.T., Luan X., Simmons D., Chen J.;
RT "Identification of a novel isoform of mouse dentin matrix protein 1:
spatial expression in mineralized tissues."
J. Bone Miner. Res. 13:422-431(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN-129/SVJ; TISSUE-Spleen;
RC Feng J.Q., Traianedes K., Luan X., McDougall M.;
RT "Study of murine Dmp-1 gene function and regulation."
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR
MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN
ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE
PARTICULARLY IN OSTEOBLAST.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; U65020; AAB93764.1;
CC EMBL; AJ242625; CAB59629.1;
CC MGI; MGI:94910; Dmp1.
CC Extracellular matrix; Signal.
CC SIGNAL 1 503 POTENTIAL.
CC CHAIN 17 503 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.

```

cervisiae chromosome XI contains the UBI2 and MPL1 genes and three
new open reading frames."
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: NOT KNOWN; WEAK SUPPRESSOR OF A MUTANT OF THE
CC SUBUNIT AC40 OF DNA DEPENDANT RNA POLYMERASE I AND III.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: L11275; AAA35091.1; -
DR EMBL; X73541; CAA51946.1; -
DR EMBL; Z28317; CAA82171.1; -
DR PIR; S38170; S38170.
DR SGD; S0001800; SRP40.
FT DOMAIN 25 314 ASP/SER-RICH
FT CONFLICT 400 400 G -> N (IN REF. 1).
FT SEQUENCE 406 AA; 41015 MW; 8EA007695AF4BAID CRC64;
SQ

```

[illegible]

RESULT 4
 MYL YEAST
 D AMYH_YEAST STANDARD; PRT; 1367 AA.
 E08640; P08068;
 01-AUG-1988 (Rel. 08, Created)
 01-FEB-1995 (Rel. 31, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
 glucosidase) (1,4-alpha-D-glucan glucohydrolase).
 STAL OR STAZ OR MAL5 OR YIR019C.
 Saccharomyces cerevisiae (Baker's yeast).
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=S288C / AB972;

car40 OR YK092C OR YKR412A.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
STRAIN-ATCC 2583 / FL100;
Lalo D., Carles C., Sentenac A., Thuriaux P.;
Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
MEDLINE=9420265; PubMed=8154186;
Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
"The complete sequence of a 15,820 bp segment of Saccharomyces

Wed Mar 26 09:38:55 2003

RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D., Churche C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A., Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M., Louis E., Iye G., Moule S., Mouton T., Odell C., Pearson D., Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V., Walsh S.V., Whitehead S.,

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

RX MEDLINE=87194600; PubMed=3106330;

RA Yamashita I., Nakamura M., Fukui S.;

RT "Gene fusion is a possible mechanism underlying the evolution of

RT STAL.";

RL J. Bacteriol. 169:2142-2149 (1987).

RN [3]

RP SEQUENCE OF 1-31 FROM N.A.

RX STRAIN-SPX101-1C; PubMed=3141213;

RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;

RT "Similar short elements in the 5' regions of the STA2 and SGA genes

RT from Saccharomyces cerevisiae.";

RL FEBS Lett. 239:179-184 (1988).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -1- SIMILARITY: TO S.POMBE SPC215.13.

CC -1- SIMILARITY: SOME. TO S.POMBE SPC285.13C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; Z38061; CAA86176.1; -

DR EMBL; M16164; AAA35014.1; -

DR EMBL; M16165; AAA35015.1; -

DR EMBL; X13857; CAA32069.1; -

DR PIR; B26877; B26877.

DR PIR; A26877; A26877.

DR PIR; S48478; S48478.

DR SGD; S0001458; MUC1.

DR Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;

KW Signal; Multigene family.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.

FT DOMAIN 210 1367 SER/THR-RICH.

FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BD61AA9D CRC64;

Query Match 4.2%; Score 190.5; DB 1; Length 1367;

Best Local Similarity 22.3%; Pred. No. 0.021;

Matches 124; Conservative 64; Mismatches 270; Indels 99; Gaps 18;

QY 151 NNDPVTFLSCGSDGVNRFDFTRIKTSCTKEDKDDLLNCRAATSAVCAIPPIYLAAG 210

DB 272 PPPPTTSTCKEPPPHDHT---TPCFKK---KTTSTKCTKTKTT-----PVPTPSSST 321

QY 211 CSDSSVRIYDRMLTRATGAGVGTGVMARFIP-----SHLNKSKCVTSIACVSEDGQ 266

DB 322 TESSAPVPTPSSSTTESSAPVPTSTTESSAPVPTPSSSTTESSAPVPTSTTESSA 381

QY 267 EILVSSSDIYLEDKDDTARELTPSAERREELRPPVKRLRGWSDTGPRAPE 326

DB 382 PVTSST-----ESSAPVPTPS-----STTESSAPV 410

QY 327 SERERDGEQSPNVLMSQMSDMLSRWFEEASEVAQSNCRGSRP-RGQTSQDSTLTPT 385

DB 411 TSSTTESSAP-----VTSSTTESSAPVPTSTTESSAPVPTSTTESSAPVPT 460

QY 386 VPSPDPLENSE---TAMEVDTPAEQFLQPSSTSSMAQAHSSTSPTES---PHSTPLLS 439

DB 461 -PSSSTTESSAPVPTSTTESSAPVPTPSSSTTESSAPVPTSTTESSAPVPTPSSST 519

QY 440 PDSQRQSVASGHHTHQS-----DNNNEKLS---PKPGTGEVPLSLHYSTEGTTSTIK 492

DB 520 TESSAPAPVPTPSSSTTESSAPVPTSTTESSAPVPTPSSSTTESSAPVPTSTTESSA 579

QY 493 LNFTDEWSSIASRRGIGSHCKSGQESF---VPQSSVOPPEGDSETKAPESESSEVDVK 549

DB 580 PVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTSTTESSAPV 639

QY 550 YQEGVSAEN---PVENHINIQTQDKFTAKPLDSNGSGERNDLNRSCGVPESASSEKAK 606

DB 640 TPSSSTTESSAPVPTPSSSTTESSAPVPTPSS-----TTTESSAPVPTSTTESS 692

QY 607 EPETSDQSTESA-----TNENNTNPEQFOFOTEA-----GPSAHEETSTRDSA 650

DB 693 APVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 752

QY 651 LQDTDDSDDDPVLIPGA 667

DB 753 TSSTTESSAPVPTPSS 769

RESULT 5

MGC1_HUMAN STANDARD; PRT; 1142 AA.

ID MGC1_HUMAN

AC G60732; O75451;

DT 15-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 13-JUN-2002 (Rel. 41, Last annotation update)

DE Melanoma-associated antigen C1 (MAGE-C1 antigen) (Cancer-testis

DE antigen CT7).

GN MAGEC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RP [1]

RN SEQUENCE FROM N.A.

RC TISSUE=Melanoma;

RX MEDLINE=98143561; PubMed=9485030;

RA Lucas S., De Smet C., Arden K.C., Viars C.S., Lethe B., Lurquin C.,

RA Boon T.;

RT "Identification of a new MAGE gene with tumor-specific expression by

RT representational difference analysis.";

RL Cancer Res. 58:743-752(1998).

RN [2]

RP SEQUENCE FROM N.A.

RX TISSUE=Melanoma;

RX MEDLINE=98284033; PubMed=9618514;

RA Chen Y.-T., Gure A.O., Tsang S., Stockert E., Jager E., Knuth A.,

RA Old L.J.;

RT "Identification of multiple cancer/testis antigens by allogeneic

RT antibody screening of a melanoma cell line library.";

RL Proc. Natl. Acad. Sci. U.S.A. 95:6919-6923(1998).

RN [3]

RP SEQUENCE FROM N.A.

RA Bird C., Heath P.;

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

CC -1- TISSUE SPECIFICITY: Expressed in testis and in tumors of a wide

CC variety of histologic types.

CC -1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

```

AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGAL
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=9223499; Pubmed=1373717;
RA Poulsen K, Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.",
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain pro-|-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87489; AAA24966.1; --
DR MEROPS; S06.001; --
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertact_sup.
DR Pfam; PF02395; IGAL; 1.
DR Pfam; PF03212; Pertactin; 2.
DR PRINTS; PR00921; IGASERPTASE.
DR Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT FT K.
FT REPEAT 1109 1116 1.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match
Best Local Similarity 4.2%; Score 188; DB 1; Length 1702;
Matches 150; Conservative 120; Mismatches 311; Indels 212; Gaps 33;

JOY 67 LSGSDDTKLIVSNPYSRKVLTIIRSHGRANIFSAKFLPCTNDKQIVSCSGDGVTFYFNVE 126
DB 773 ITASDNKAVHIG--YKAGDTVCVRSDDYG-----YVCTTDK--LSDKALNSFNATNVS 822
JOY 127 ODAETNRCQCFCH-----YGR-----IYEIMVPNDPYTFLSCGEDGTVRWFOTRIKTS 177
DB 823 GNVNLSGGNANFVLGKANLFGIISTGNSQVRLTNSHWLGTG---DSNVQLNLD----- 874
JOY 178 TKEDCKDILINCRRATSAVACPPIP-----YILAVGCDSSSVRIYDRRMLGTR 227
DB 875 -----KGHIHLNAQNDANKVTYNTLTVNSLSGNGSFYLL-----TDLNKGQGDVVVTKS 925

```

Wed Mar 26 09:38:55 2003

QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSKRVSLCYSED---GQELVSYSDYI 277
 Db 926 ATGNTLVADTKTGETPKNELTLFDASNATRNLLNSLVGNTVDLGAWKYKLRNNGRY- 984
 QY 278 YLFDPK-----DTPAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318
 Db 985 DLYNPEVEKRNQTDVDTNITPNNTQADVPSPNSNEELIARVETVPVPPAPATPSETTET 1044
 QY 319 TGPAPRES-----ERERDQ-----EQSPNVSLMQMSDMLSWFE-----354
 Db 1045 VANSQKSEKTEVKNEDATETTTAQNAGEVAEAKPSKANTQNTNEAQSGETEETOTTE 1104
 QY 355 --BASEVAQNRGRSRPRGTSQSDISLTPVPS-----SPDLEVSETA 398
 Db 1105 IKETAKVEEKAKEEKAKEEKAKEEKAKEEKAKEEKAKEEKAKEEKAKEEKAKEEKAKEE 1164
 QY 399 MEVDTPAEQFLOPSTSTMSAQAHSTSPSTESPHSTPLSSPDSEQRQSVASGHHTHQ 458
 Db 1165 VQA-----QPQTSTVAAAEAT-----SPNKPAAETQPSKTAEPVTPVVKKNQ 1211
 QY 459 SDNNEKLSKPRGTPGEPVLSHYSTEGTSTIKLNFDEWSSSIASSRGIGSHCKSE 518
 Db 1212 TENT-----TDQP-----TEREKAKVETETKTPQPVASQA-----SPKQEQ 1249
 QY 519 ESFVQSSVOPPEGDSSTKAPESSESDVTKYQGVSAENPV-ENHIN-----ITQSDK 571
 Db 1250 SETVQPAVLSENVPTVYNAEEVQAQLQTSATVSTKQAPENSINTGSATAITAE 1309
 QY 572 FTAKPLDSNGERDNLDRSCGVPEESASSEKAKEPSTSDQSTESATENNTNPEPF 631
 Db 1310 KSDKP-----QTETAATEDASQHKANTVADNSVANNSESSEPKSRR 1351
 QY 632 QTEATPSAHEETSTSDSALQDSDSDVLPICARVAGRGDRRSARVARIQEFFRRK 691
 Db 1352 RRSISQ---QETSAEETTAADTET-----TIADNSKRKSP-NRRS-----RRS 1392
 QY 692 ERKEMBELDTLNRRLPLVMYKGRHSRTMLKEANFWGNCVMSGDCGHIFI---WDR 748
 Db 1393 VRSE-----PTVTVNGSRSTVALRDLTSTNTNAVISDAMAKAQFVALNVGK 1438
 QY 749 HTAEHLMLLEADN 761
 Db 1439 AVSGHISQLEMMN 1451
 RESULT 7
 IGAO_HAEIN STANDARD; PRT; 1694 AA.
 AC P44969;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
 GN IGA OR IGAI OR HI0990.
 OS Haemophilus influenzae.
 OC Bacteria; proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Serotype D;
 RA Wright A., Fishman Y., Tai F., Plaut A.G.;
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
 CC certain Pro|-xaa bonds in the hinge region. No small molecule
 CC substrates are known.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X59800; ; NOT_ANNOTATED_CDS.
 CC EMBL; U32779; AAC22651.1; ;
 CC MEROPS: S06.001; ;
 CC TIGR: HI0990; ;
 CC InterPro: IPR000710; IGA_S6.
 CC Pfam: PF023395; IGAI; 1
 CC Pfam: PF03212; Pertactin; 2
 CC PRINTS: PR00921; IGASERPTASE.
 CC HydroLase; Serine protease; Transmembrane; Zymogen; Signal;
 CC Complete proteome.
 CC SIGNAL 25
 CC CHAIN 1 26 1014
 CC PROPEP 1015 1694
 CC FT ACT_SITE 288
 CC FT CONFLICT 253
 CC FT CONFLICT 272
 CC FT CONFLICT 464
 CC FT CONFLICT 866
 CC FT CONFLICT 1036
 CC FT CONFLICT 1074
 CC FT CONFLICT 1421
 CC FT CONFLICT 1545
 CC FT CONFLICT 1694 AA; 185539 MW; C52427013F93178C CRC64;
 SQ
 Query Match 4.1%; Score 186; DB 1; Length 1694;
 Best Local Similarity 19.4%; Pred. No. 0.045;
 Matches 153; Conservative 113; Mismatches 310; Indels 214; Gaps 33;
 QY 67 LSGSDDTKLVISNPYSRKVLTITIRSGHRANIFSAKFLPCTNDKQIVSCSDGVIFVTNVE 126
 Db 773 ITASDNKAVHIG--YKAGDTVCVRSDYTG-----YVCTCTDK--LSDKALNSFNATNVS 822
 QY 127 QDAETNRQCOFTCH-----YGT-----TYELMTVPNDPYFLSCGEGDTVRWFDTRIKTSC 177
 Db 823 GNVNLSGNANFVLGKANLFGTISGTGNSQVRLTENSHTLTG---DSNVNQLND-----874
 QY 178 TKEDCKDDILLNCRRRAATSVACPPIP-----YLLAVGSDSSSVRIYDRMLGTR 227
 Db 875 -----KGHIHLNAQNDANKVVTYNTLTAVNSLSGNGSFYLL-----TDLNKGQGVVTKS 925
 QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSKRVSLCYSED---GQELVSYSDYI 277
 Db 926 ATGNTLVADTKTGETPKNELTLFDASNATRNLLNSLVGNTVDLGAWKYKLRNNGRY- 984
 QY 278 YLFDPK-----DTPAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318

Db 985 DYNPEVERKNOTVDITNTITNNIQADVPSPSNNEETARVETPPPPAPAPSETTET 1044
 QY 319 TGPAPPEREREDGQSPVNSLMORMSDMLSRWFEEA-----SEVAQS----- 362
 Db 1045 VAENSKOESKTEKNEQDATTAAQ-----GEVAEEAKPSVKANTQTNEVAQSGTTEE 1099
 QY 363 -----NRGRSRPRGTSOSDITLTPVPS-----SPDLEVSSTAMEV 401
 Db 1100 TOTTEIKETAKVEKEKVEKDEIQEAPOMASETSPKQAKPAKEVSTDDIKVEETQVOA 1159
 QY 402 DTPAQFLOPSTSSMSAQAHSTSPSPHSTPLLSPPDSEORQSVBASGHHTTHQSDN 461
 Db 1160 -----QPQTSITVAAAEAT-----SPNSKPAEETQPSSEKNAEPVTPVSKNQEN 1206
 QY 462 NNEKLSKPGGEPVLSHYSTEGTITIKLFTDEWSSIASSSRGIGSHCKSEGQES 521
 Db 1207 T-----TDOP-----TEREKTAKVETKTQPPQVASOA-----SPKOEQSET 1244
 QY 522 FVPOSSVQPPGQSETKAPRESSDVTYQGVSAENPV-ENHIN-----ITQSDKFTA 574
 Db 1245 VQPAVLESENVPTVNAEVEQALQOTQTSATVSTKQAPENSINTGSATAITETAESD 1304
 QY 575 KPLDSNGERDNLDRSCGVPEESASSEKAKPETSDQTSATNTNENPQOTE 634
 Db 1305 KP-----QTETAATEDASQHKANTVADNSVANNSESSDPKSRRRS 1346
 QY 635 ATGPSAHEETSTRSALQDQDDPVLIPGARYAGDRSAYARIOEFFRRK 694
 Db 1347 ISQP-----QETSARETTAASTDET-----TIADNSKRSKP-NRRS-----RSVR 1387
 QY 695 EMBELTLNIRPLVKNVYKGRHSRTMIKEANFWANFVMSGDCGHFI---WDRHTA 751
 Db 1388 E-----PTVNGSDRSTVALRDLTSTINAVISDAMAKAQFVALNVGKAYS 1433
 QY 752 EHLMLLEADN 761
 Db 1434 QHISOLEMNN 1443

RESULT 8

HIRA_FUGRU
 ID HIRA_FUGRU STANDARD; PRT: 1025 AA.
 AC 042611;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE HIRA protein (TUPI like enhancer of split protein 1).
 GN HIRA OR TUPI1.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98201624; PubMed=9524281;
 RA Llevadot R., Estivill X., Scambler P., Pritchard M.;
 RT "Isolation and genomic characterization of the TUPI1/HIRA gene of
 RL the pufferfish Fugu rubripes.";
 RL Gene 208:279-283(1998).
 CC -!- FUNCTION: Could have a part in mechanisms of transcriptional
 CC regulation similar to that played by yeast HIR1 and HIR2 together.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
 CC -!- SIMILARITY: BELONGS TO THE HIRA FAMILY OF WD-REPEAT PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U94325; AAC60370.1; -
 DR EMBL; U94324; AAC60369.1; -
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 8
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 3.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS00082; WD_REPEATS_2; 3.
 DR PROSITE; PS02094; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repeat; WD repeat; Nuclear protein.
 FT REPEAT 68 98
 FT REPEAT 129 159
 FT REPEAT 172 202
 FT REPEAT 266 313
 FT DOMAIN 673 682
 FT DOMAIN 685 688
 FT POLY-ALA.
 SQ SEQUENCE 1025 AA; 111856 MW; A4212152D75B6A37 CRC64;
 Query Match 4.13; Score 184; DB 1; Length 1025;
 Best Local Similarity 20.08; Pred. No. 0.031;
 Matches 175; Conservative 116; Mismatches 312; Indels 272; Gaps 39;
 QY 42 KLEATLVNHDGCVNTICWNTDTEVILSGSDTKLV-----ISNPY 81
 Db 61 KMLCDMNHLCACVNCVNSNGLYLASGGD-KLVVWKRAALIGPSTVFSSKNLANVE 119
 QY 82 SRKVLATIRSHGRANTFSAKFLPCTNDKQIVSCSGDGVFTVNVQDAETNRQCFCHY 141
 Db 120 QWRCVTLIRN-HTGDMVDSWSP--HDVWLASGVSDNTIIVNARKEPE-----MTCLR 171
 QY 142 GTTYEIMTVNDPY--TFLSCGEDGTVR-W--FDTRIKTSCTKEDCKDDILINCRRAATS 196
 Db 172 GHTGLVKGLTWDPVGKYIASQADDSLRVWRTVDMQMEANITKPFSECGGTTHVLRSWS 231
 QY 197 VAICPPIPYVL-----AVGSDSSVRIYDRMLGTRATGNAGRTTGMVARETPSHLNK 252
 Db 232 -----PDGQYLVSAHAMNNSGPTAQIVERD--GWRTNMDVFGHRKAVTVVKNPKFKKK 284
 QY 253 -----SC-----RVTSICYSEDEQEL 269
 Db 285 QKNGGSPKPCPCVCCAVGSKDRSLVMTSLKRLPLWIHDLFKSIMDISWTLTGLGML 344
 QY 270 V-SYSSDIYILDPKDDTARELKTPSAEERRELRQ-----P 305
 Db 345 VCSMDGTVAYL-----DFSLDELGDPLSEEEKNSIQNIYVKSIAITNTEPQLSTIENP 400
 QY 306 PV---KRLRLRGDWSDTGP-----RARPESERERDQSPVNSLMQMSDMLSRWFEEA 356
 Db 401 EMLKYQQERRNSTQANSQPGATGESATPKLNSVMNGE-----SLEDIRKNLLKQVE-- 453
 QY 357 SEVAQSNRGRSRP-----RGTSQSDISTLTPVPSDPLEVSEAMEVDPAEQL 409
 Db 454 ---TRTPDGRRIITPLCIAQLDGTDFSPALFNSAPILPSG----- 490
 QY 410 QPSTSTMSAQAHSTSPSTES-----PHSTPILLSPPDSEORQSVASGHHTTHQSDNNN 463
 Db 491 -SSMSNQLTSQSSDSSPGQAPPLGLRPSQDPMILISPPSSAAKV-----LEDNKD 540
 QY 464 -----EKLSPKPGTGPVLSHYSTEGTITIKLFTDEWSSIASSSRGIGSH 512
 Db 541 GVKSCILLTSASKIEPMKALDSRTFSKATPGAT-----AAIASST-GL-- 584
 QY 513 CKSEGESEFVPOSSVQPPPE--GDSETK-----APE-----ESSEVDVTKYOE 553
 Db 585 TPSEPKETPKQDKVEDTSSDESKMATINKNLAFNKRKPELLMDGAEEVKKKG 644
 QY 554 VSAENPVENHINTQSDKFTAKPL--DSNSGENDLNLDRSCGVPEESASSEKAKPET 611
 Db 645 RPRKDMAASI-----AQPLTQTTSPAEEP---SRAAAGAGAAAPTAAAKLP 692

Wed Mar 26 09:38:55 2003

```

QY 612 DOTSTESATNTNTPQFOFTEATGSAHEETS-----TRDSALQDTDDSDPVLIPGA 667
Db 693 TFSIKKAFQVSMDFVSVLEVE-----NEVSVAGSRUSQLRCSRDGRDWTLLPSS 745
QY 668 RYRAGPDRSVA---RIQEFRRKKEKEMEELDTLIRPLVKMKYKGRHSRTWIK 724
Db 746 VITAAGSDVAVASQDRMLSVFSCGRL-----LPAIQLATP-----A 785
QY 725 EAFNFWANFVMSGDCGHFIWDRHTAHLMLLEA 759
Db 786 SALHCSAHEFVWLTSATLSVMDVHKQKALVKNES 820

RESULT 9
YMF7_YEAST STANDARD; PRT: 1658 AA.
ID YMF7_YEAST
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUA1-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
SEQUENCE OF 1-711 FROM N.A.
RP STRAIN=S288c / AB972;
RC Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 608-1648 FROM N.A.
RP STRAIN=S288c / AB972;
RC Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[3]
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; 249809; CAA89934.1; -
CC EMBL; 249939; CAA90190.1; -
CC SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.9%; Score 177.5; DB 1; Length 1658;
Best Local Similarity 18.1%; Pred. No. 0.12;
Matches 116; Conservative 110; Mismatches 243; Indels 171; Gaps 25;

QY 278 YLEPDKDITARELTPSAEERREL--RQPPVKRLRLRGDWSGTGPRPARESERDEGEQ 335
Db 169 FISSPLPADESSNIDSDKDELEGKQSLIKDFLENDYEYLS-----EEKNSDQGS 223
QY 336 SPNVSL-----NORMSDMLSRWFEEASVAGNSRGRSRPRGTT--SQSDIST 382
Db 224 SPSMILSDEYAEAGALQVSN--DEYAEEGQVGRKNIGQEQANVENATQISSDSE 281
QY 383 LPTVPSPDLEVSAMEVDTPAE-----QFLQPTSTSTMSQAHS 425
Db 282 QONYSEGEMEL-EDDIDVESDAEKDSQAGTEHSVDFSKYMPQRPDNT----- 331
QY 426 SPTEPHSTPLSSPDSEQRQSVASGHHTTHDSNNNEKLSPKPGTGPVLSLHYSTEG 485
Db 332 -----KIPVIEKVESD-----HKVHQ-----RYSBDG 354

```

RESULT 10

IGM4_HAEIN STANDARD; PRT: 1849 AA.
ID IGM4_HAEIN
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxID=727;
RN [1]
RP STRAIN=NETI HK61;
RC MEDLINE=92234949; PubMed=1373717;
RX Poulson K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M87491; AAA24968.1;
CC MEROPS; S06.001; -
DR

DR InterPro: IPR000710; Iga_s6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactid; 2.
DR PRINTS: PR00921; IGASERTPASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021
FT PROPEP 1022 1849
FT ACT_SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 3.9%; Score 175.5; DB 1; Length 1849;
Best Local Similarity 21.6%; Pred. No. 0.16;
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;

QY 250 NKKSRV-----TSLCYSDGGQILVSYSDVLYLFDPKDDTARELKTPTSAEERR 299
Db 921 NKKSKVNVVNSARGNETLQVADKTGP-----NHNELTLDASNATRNNEVLTANGSV 975
QY 300 EELRQPPVKRLRLGDNS----DTGPR---ARPESERRDGEQSPNVSLMQMSDMLSRW 352
Db 976 D-----RGANKYKLRNVNGRYDLYNVEVEKRNQTVDTNITTPNDI----- 1016
QY 353 FEEASEVAQNRGRSRPRGGTSQSDISTL--PTVPSSPDLEVSETAMEVDTPAEQFLQ 410
Db 1017 -QADAPSAQSN-----NEEIARVETPVPPAPATESAIAAGQETRAETAQA 1062
QY 411 P-----STSTWSAQAHSTSPSPHPTLLSSPDSQSQSVASGHHTHHOSDNNNE 464
Db 1063 PAMEETNTANSTAPKSDATQTQTNPNSE--SVSPETTEKVAENPQENETVAKNEQE 1119
QY 465 KLSPKPGTGB-----PVLISHYSTEGTTTSTIKLNFEDWSSTASSRGISGCKSEGG 518
Db 1120 ATPTPQNGEAKEDOPTVEANTQNEATQSGK---TEETQATSEPTSESVTSVSENG 1176
QY 519 EESFVQSSQVPPGDSSEKAPESSEVDYTKQGVSAENPVENHINTOSDKFTAK-PL 577
Db 1177 PEKTVSQSTEDKVVVEKEKAKVE--TEETQAPQVTSKEPP-----KQAEPAPEVPT 1228
QY 578 DNSGGERNDLNRDCGVPESASSEKAKEPETSDQSTESATNENNTPEPQFTEATG 637
Db 1229 DTNAEEAALQOQTPTTVAEAETSPNSKPAETQOPS-----EKTNAEP-----VT 1275
QY 638 PSAHEETSTRSALQDTSDDPVL-IPGARVRAGPGDRRSARVARIQEFFFFRRKKERKM 696
Db 1276 PVVSENATQPTETETAKVEKTEQVPOVASQESPKQBQPA-AKPAQTKPQAEPAPE 1334
QY 697 EELDTLNRPP 707
Db 1335 NVLTKNVGP 1345

RESULT 11
DPPP_MOUSE STANDARD; PRT; 934 AA.
AC P97399; Q70567;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dentin sialoprotein precursor (Dentin matrix protein-3) (DMP-3)
DE 3) [Contains: Dentin phosphoprotein (Dentin phosphoprotein) (DPP);
DE Dentin sialoprotein (DSP)].
GN DPPP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RC STRAIN-Swiss Webster; TISSUE=Molar;
RA MEDLINE=97150835; PubMed=8995371;
RA MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;

"Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. Dentin phosphoprotein DNA sequence determination."; J. Biol. Chem. 272:835-842(1997).
[2]
SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS. STRAIN=129/SVJ; TISSUE=Liver; MEDLINE=98211969; PubMed=9545272; Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B., D'Souza R.N., Kozak C.A., MacDougall M.; "Genomic organization, chromosomal mapping, and promoter analysis of the mouse dentin sialoprotein (Dsp) gene, which codes for both dentin sialoprotein and dentin phosphoprotein."; J. Biol. Chem. 273:9457-9464(1998).
[3]
SEQUENCE FROM N.A. STRAIN=129/SVJ; TISSUE=Liver; Steir C., Butler S., Lin E., George A., Veis A.; "From mouse to zebrafish-dentin matrix proteins genomic characterization."; Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[4]
TISSUE SPECIFICITY. MEDLINE=98055479; PubMed=9395101; Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L., Hotten D., Papageorgakis P., Berdal A., Butler W.T.; "Dentin sialoprotein (DSP) transcripts: developmentally-sustained expression in odontoblasts and transient expression in pre-ameloblasts"; Eur. J. Oral Sci. 105:405-413(1997).
[5]
TISSUE SPECIFICITY. MEDLINE=21096982; PubMed=11175790; Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M., Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z., Zhao G., Kong X.; "Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP."; Nat. Genet. 27:201-204(2001).
-!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP may bind high amount of calcium and facilitate initial mineralization of dentin matrix collagen as well as regulate the size and shape of the crystals.
-!- SUBCELLULAR LOCATION: Secreted (By similarity).
-!- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and transiently in pre-ameloblasts. Found in the inner ear.
-!- PFM: DSP is glycosylated.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to license@isb-sib.ch.

EMBL; U67916; AAC12787.1; .
EMBL; AJ002141; CA05208.1; ALT_SEQ.
DR MGD; MGI:109172; Dsp.
KW Signal; Extracellular matrix; Glycoprotein; Sialic acid; Phosphorylation.
FT SIGNAL 1 17
FT CHAIN 18 934
FT CHAIN 18 451
FT CHAIN 452 934
FT DOMAIN 419 934
FT SITE 479 481
FT MOD_RES 227 227
FT MOD_RES 254 254
FT MOD_RES 279 279
FT MOD_RES 293 293
FT MOD_RES 299 299
POTENTIAL.
DENTIN SIALOPHOSPHOPROTEIN.
DENTIN SIALOPROTEIN.
DENTIN PHOSPHOPROTEIN.
SER/ASP-RICH.
CELL ATTACHMENT SITE (POTENTIAL).
PHOSPHORYLATION (BY CK2) (POTENTIAL).
PHOSPHORYLATION (BY CK1) (POTENTIAL).
PHOSPHORYLATION (BY CK1) (POTENTIAL).
PHOSPHORYLATION (BY CK2) (BY SIMILARITY).
PHOSPHORYLATION (BY CK1) (BY SIMILARITY).

Wed Mar 26 09:38:55 2003

FT MOD_RES 314 314 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT MOD_RES 336 336 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT MOD_RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 934 934 A618789D8A57249A CRC64;
 SQ SEQUENCE 934 AA: 93901 MW; 1818789D8A57249A CRC64;
 Query Match 3.9%; Score 175; DB 1; Length 934;
 Best Local Similarity 19.6%; Pred. No. 0.079;
 Matches 84; Conservative 78; Mismatches 205; Indels 62; Gaps 9;
 QY 262 SEDQELTVSYSDYVLFDPKDDTARELKTGSAERREELRQPPVVKRLRLRGDMSDTGP 321
 Db 525 SEDKDE---SDSDH---DNSSDS--ESKSDSDSDSDS-----SDSDSDS- 562
 QY 322 RARPESEERDGGSPNVSLMQRMDSMLSNFEEASVAVQSNRGRSRPRGTSQSDIS 381
 Db 563 ---DSSDSDSDSDSDS-----DSSDSDSDSDSDSDSDSDSDSDSDSDSDS 604
 QY 382 -----TLTPVPSPLLEVSTAMEYDTTAEQFLOPSTSTMSAQAHSTSPTESPHST 434
 Db 605 DTCSDS 664
 QY 435 PLLSSPDEQSGVPAAGHTTHQSDNNKLSPKPGTGEVPLSLHYSTEGTTSTTKLN 494
 Db 665 DSSDS 724
 QY 495 FDEWSS-----IASSRGIGSHCKSEGOESFVQSQVPEGDSKAPESSEIDVTK 549
 Db 725 SDS 784
 QY 550 YEGVSAENPVENHINITQDKFTAKPLDSNGERNDLNLDRSCGVPEASAKAPE 609
 Db 785 SDS 844
 QY 610 TSQQTSTESATNENNTNPEPQFOTATGSAHEETSTRDSALQDTHDDPVLIPGARY 669
 Db 845 SDS 895
 QY 670 RAGPDERS 678
 Db 896 -SDGDSKS 903
 RESULT 12
 ID NASP_HUMAN STANDARD; PRT; 787 AA.
 AC P49321; 1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-FEB-1996 (Rel. 41, Last annotation update)
 DE Nuclear autoantigenic sperm protein (NASP).
 GN NASP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=93050782; PubMed=1426632;
 RA O'Rand M.G., Richardson R.T., Zimmerman L.J., Widgren E.E.;
 RT "Sequence and localization of human NASP: conservation of a xenopus
 histone-binding protein.";
 RL Dev. Biol. 154:37-44 (1992).
 CC -1- FUNCTION: MAY PLAY A ROLE IN REGULATING THE EARLY EVENTS OF
 SPERMATOGENESIS BY BINDING NEWLY SYNTHESIZED HISTONES VARIANTS AND
 TRANSPORTING THEM TO THE NUCLEUS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: TESTIS- AND SPERM-SPECIFIC PROTEIN.
 CC -1- SIMILARITY: WITH XENOPUS HISTONE-BINDING PROTEIN NL/N2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M97856; AAA36027.1;
 DR GenBank: HGNC:7644; NASP.
 DR MIM: 603185;
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 3.
 DR SMART: SM00028; TPR; 3.
 KW Nuclear protein; Sperm; Coiled coil; Antigen.
 FT DOMAIN 136 164 COILED COIL (POTENTIAL).
 FT DOMAIN 460 487 COILED COIL (POTENTIAL).
 FT DOMAIN 597 664 COILED COIL (POTENTIAL).
 FT DOMAIN 752 769 COILED COIL (POTENTIAL).
 FT DOMAIN 116 126 GLU-RICH (ACIDIC).
 FT DOMAIN 462 511 GLU-RICH (ACIDIC).
 FT DOMAIN 715 721 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT SEQUENCE 787 AA: 85130 MW; 18DBA6DFD48321E0 CRC64;
 SQ SEQUENCE 787 AA: 85130 MW; 18DBA6DFD48321E0 CRC64;
 Query Match 3.9%; Score 174.5; DB 1; Length 787;
 Best Local Similarity 19.5%; Pred. No. 0.068;
 Matches 126; Conservative 96; Mismatches 228; Indels 195; Gaps 29;
 QY 283 KDTARELKTGSAERREELRQPPVVKRLRLRGDMSDTGPR--ARPESEERDEQ----- 335
 Db 124 EDESLVENNDIDEAREELREQVYDAMGEKAEKTEDKSLAKPETDKEQDSEMEKGR 183
 QY 336 -----SPNVSLMQRMDSMLSNFEEASVAVQSNRGRSRPRGTSQSDISLTTPVSSP 390
 Db 184 EDMDISKSAEPEQEKVDTLTLDTETSEEA-----KGAAPE-GPNEAEVTSKPEQEV 237
 QY 391 DLE-----VSTAMEYDTTAEQFLOPSTSTMSAQAHSTSPST 428
 Db 238 DAEEKSVSGTVDQECREKGGQEGEVIVSIEKPEVSEEPVVTLEKQGTAVEVEA 297
 QY 429 RS--PHSTPLSSPDSEQRQSVASGHHTHQSDNNKLSKPKGTGEVPLSLHYSTEGT 486
 Db 298 ESIDPTVRPVDVGDEPEKVV-----TSENAEKAVLEQLVGOVPPAPEPEVQ 348
 QY 487 TTSTIKLNTDEWSSIASRSGIGSHCKSEGOESFVQ--SSVQPE--GDSEKAPES 543
 Db 349 TEAA-----EASAVEAGS-----EVSEKPGQEAFLPGAVNGPSVVGDTPIEPQTS 397
 QY 544 SEDVTYKQGVSAENPVENHINITQDKFTAKPLDSNGERNDLNLDRSCGVPEASASSE 603
 Db 398 IERLTQKDGSGLEEKV-----RAKLPSEETKL 427
 QY 604 KAKEPETS-DQTSATNENNTNPEPQFOTATGSAHEETSTRDSAL---QDTHDSD 659
 Db 428 SVESEAGDGVDTKVAQGAATEKSPEDKVQI-----AAEETQEREQKMEETEGSEE 482
 QY 660 DPVLIPGARYRAGPDERSAVARIQEFRRRKRERKEEMELDITLNI---RRLPVKVKYKH 716
 Db 483 DD-----KENDTEEMPNDVSLNKSLOENEEIEIGLEAWDMLDLAKIIFK-- 530
 QY 717 RNSRTMIKEANFWANFVMSGDCGHIFWDRHTAEHML---LEADNHV-----VNC 766
 Db 531 ---RQETKEALYAAQ-----AHLKLGVSVESENYQVAVEEFQSC 568
 QY 767 -----LQHPFPILASS-----GIDYDIKWSPLSESRIFNKKLADEVITRNLML 815
 Db 569 LNLQEQYLEAH--DRLLAEHYQLGLAYG-----YNSQY-DEAVAQFSKSIEV 613
 QY 816 TRNTITVPASFLMLRLASLNHIRADRLDGDSESGSQENENEDEE. 860

[illegible]

Wed Mar 26 09:38:55 2003

- RA Colleaux L., Schwartz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding
 RL a potential zinc finger helixase.";
 RN Genomics 43:149-155(1997).
 [3]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE-95179111; PubMed-7874112;
 RA Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,
 RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helixase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 [4]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE-94214473; PubMed-8162050;
 RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,
 RA Milasseau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 [5]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE-95211835; PubMed-7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RX Pearce A., Chapman J.;
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 [7]
 RP EZH2 BINDING.
 RX MEDLINE-98167853; PubMed-9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATRX gene product and the SET
 RT domain of the human EZH2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE-20040663; PubMed-10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Blackmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 [9]
 RP MEDLINE-20213147; PubMed-10751095; Gecz J.;
 RX Villard L., Fontes M., Ades L.C.;
 RA "Identification of a mutation in the XNP/ATRX gene in a family
 RA reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE-97196774; PubMed-9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RA "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 [11]
 RP VARIANT JM GLN-2131.
 RX MEDLINE-96224392; PubMed-8630485;
 RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugter-Weber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 [12]
 RP VARIANTS ATR-X.
 RX MEDLINE-97467722; PubMed-93269311;
 RA Gibbons R.J., Bachoo S.A., Picketts D.J., Aftimos S., Aisenbauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 [13]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE-20123062; PubMed-10660327;
 RA Fichera M., Romano C., Castiglia L., Pailla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RA "New mutations in XNP/ATRX gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE-99347960; PubMed-10417298;
 RA Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RA "Mutation of the XNP/ATRX gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 [15]
 RP VARIANT CWS THR-2050.
 RX MEDLINE-99326061; PubMed-10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 [16]
 RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE-99219535; PubMed-10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belouigne J.,
 RA Lissi A.-M., Seaver L., Nefont J.-P., Romano C., Fichera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 [17]
 RP VARIANTS ATR-X S-179; L-190; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE-20451413; PubMed-10995512;
 RA Wada T., Kubota T., Fukushima Y., Saitoh S.;
 RA "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 CC INTERACTING WITH HP1.
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1, 2, 3, 4 (SHOWN HERE) AND 5;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBICUITOUS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 CC SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME

[illegible]

DB 182 DS1PDSSEHHVGGGSEGDSSHGDSSEFDDGMSDDPGAYRSEGRNSRISDAGLKSTQ 241

Wed Mar 26 09:38:55 2003

DR EMBL; U89012; AAC51332.1; -
 DR EMBL; U34037; AAA97602.1; -
 DR EMBL; U65378; AAB87728.1; -
 DR Genew; HGNC:2932; DMP1.
 DR MIM; 600980; -
 KW Extracellular matrix; Signal; Alternative splicing.
 FT SIGNAL 1 16
 FT CHAIN 17 513
 FT CARBOHYD 25 25
 FT CARBOHYD 285 285
 FT CARBOHYD 324 324
 FT CARBOHYD 345 345
 FT CARBOHYD 351 351
 FT CARBOHYD 413 413
 FT CARBOHYD 426 426
 FT CARBOHYD 467 467
 FT SITE 364 366
 FT VARSPLIC 46 61
 FT CONFLICT 69 69
 FT SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CRC64;
 SQ

Query Match 3.8%; Score 171.5; DB 1; Length 513;
 Best Local Similarity 21.8%; Pred. No. 0.059;
 Matches 102; Conservative 58; Mismatches 210; Indels 97; Gaps 18;
 QY 262 SEDQELIVSYSDYIYLF-----DPKDDTARELTPSAEERREELRQPPVKR 309
 Db 69 STQSEGLGSDHDIYIYLAGGFSRSTGKGGDDDD-----EDSGDD----- 112
 QY 310 LRLRGDWSDTGPRARPERERERQEQSPNVSLMQMSDMLSRWFEEASEVAQ-----SNR 364
 Db 113 -TFGDDSGPGKDRQEGGNSRLGSDSDDTIQAESAPQGDQAQDTSESRELDNE 171
 QY 365 GGRSRPRGGTS-QSDISTLTPTVFPSPD-----LEVSETAMEVDTPAEQFLQPTSS 415
 Db 172 DRVDSKPEGSDTQSESEEHVWGGSGSHGSGELDDGQSDDP-ESISERGSNS 230
 QY 416 TMSAQA-----HSTSPSPESHPTLLSPD-----SEQRQSVFASGHHTHQ 458
 Db 231 RNSAGMKSKESGENSEQANTQDGGSGQLLEHPSRKIFRKRISDEDDRDELDDNTMEE 290
 QY 459 --SDNNKSLSPKPGTGEVPLSLHYSTEGTTTTLKNTDEM-----SSIASSSRGI 509
 Db 291 VKSDSTENSRTDGLSQP-----RRDSKGDQSDSKENLSQESQNVDPSSSEQANL 346
 QY 510 GSHCKSGQESFVQSSVQPP-----EGDSETKAPESEEDVTKQBGVSAENPV-- 560
 Db 347 SSOENSESQEEVVSERSGNDPPTTSYVEDQEDSDSEEDSSHTLSHKSSEEEQADS 406
 QY 561 ENHINITQSKFTAKPLDSDNSGERNDLNRSCGVPEASASSEKAKEPETSD-QTSTESA 619
 Db 407 ESSELSNFSPESESPEDENSQEGLOSHSSA---ESQSESHSEFDDSDSDSQRSRK 463
 QY 620 TNENNTPEPQFQTEATGSPSAHEETSTR-----DSALQDTDDSD 658
 Db 464 EDSNTE--ESKSSSEEDQQLANIESRKLTVDAYHNKPIGDQDDND 508

RESULT 17
 DEXTLSTRDO STANDARD; PRT; 1337 AA.
 ID DEXTLSTRDO
 AC P39653; 1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-glucanohydrolase).
 GN DEX.
 OS Streptococcus downei (Streptococcus sobrinus).
 OG Plasmid pYA902
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=1317;

DR EMBL; U89012; AAC51332.1; -
 DR EMBL; U34037; AAA97602.1; -
 DR EMBL; U65378; AAB87728.1; -
 DR Genew; HGNC:2932; DMP1.
 DR MIM; 600980; -
 KW Extracellular matrix; Signal; Alternative splicing.
 FT SIGNAL 1 16
 FT CHAIN 17 513
 FT CARBOHYD 25 25
 FT CARBOHYD 285 285
 FT CARBOHYD 324 324
 FT CARBOHYD 345 345
 FT CARBOHYD 351 351
 FT CARBOHYD 413 413
 FT CARBOHYD 426 426
 FT CARBOHYD 467 467
 FT SITE 364 366
 FT VARSPLIC 46 61
 FT CONFLICT 69 69
 FT SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CRC64;
 SQ

Query Match 3.8%; Score 171.5; DB 1; Length 513;
 Best Local Similarity 21.8%; Pred. No. 0.059;
 Matches 102; Conservative 58; Mismatches 210; Indels 97; Gaps 18;
 QY 262 SEDQELIVSYSDYIYLF-----DPKDDTARELTPSAEERREELRQPPVKR 309
 Db 69 STQSEGLGSDHDIYIYLAGGFSRSTGKGGDDDD-----EDSGDD----- 112
 QY 310 LRLRGDWSDTGPRARPERERERQEQSPNVSLMQMSDMLSRWFEEASEVAQ-----SNR 364
 Db 113 -TFGDDSGPGKDRQEGGNSRLGSDSDDTIQAESAPQGDQAQDTSESRELDNE 171
 QY 365 GGRSRPRGGTS-QSDISTLTPTVFPSPD-----LEVSETAMEVDTPAEQFLQPTSS 415
 Db 172 DRVDSKPEGSDTQSESEEHVWGGSGSHGSGELDDGQSDDP-ESISERGSNS 230
 QY 416 TMSAQA-----HSTSPSPESHPTLLSPD-----SEQRQSVFASGHHTHQ 458
 Db 231 RNSAGMKSKESGENSEQANTQDGGSGQLLEHPSRKIFRKRISDEDDRDELDDNTMEE 290
 QY 459 --SDNNKSLSPKPGTGEVPLSLHYSTEGTTTTLKNTDEM-----SSIASSSRGI 509
 Db 291 VKSDSTENSRTDGLSQP-----RRDSKGDQSDSKENLSQESQNVDPSSSEQANL 346
 QY 510 GSHCKSGQESFVQSSVQPP-----EGDSETKAPESEEDVTKQBGVSAENPV-- 560
 Db 347 SSOENSESQEEVVSERSGNDPPTTSYVEDQEDSDSEEDSSHTLSHKSSEEEQADS 406
 QY 561 ENHINITQSKFTAKPLDSDNSGERNDLNRSCGVPEASASSEKAKEPETSD-QTSTESA 619
 Db 407 ESSELSNFSPESESPEDENSQEGLOSHSSA---ESQSESHSEFDDSDSDSQRSRK 463
 QY 620 TNENNTPEPQFQTEATGSPSAHEETSTR-----DSALQDTDDSD 658
 Db 464 EDSNTE--ESKSSSEEDQQLANIESRKLTVDAYHNKPIGDQDDND 508

RESULT 16
 DMP1_HUMAN STANDARD; PRT; 513 AA.
 ID DMP1_HUMAN
 AC Q13316; Q43265; 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix protein-1) (DMP-1).
 GN DMP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Oolar;
 RA MEDLINE=97321043; PubMed=9177774;
 RA First K.L., Simmons D., Feng J., Apelin H., McDougall M.;
 RA "Elucidation of the sequence and the genomic organization of the human dentin matrix acidic phosphoprotein 1 (DMP1) gene: exclusion of the locus from a causative role in the pathogenesis of dentinogenesis imperfecta type II.";
 RA Genomics 42:38-45(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RA McDougall M., Juan X., Simmons D., Feng J.;
 RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.
 RA MEDLINE=96163890; PubMed=8586437;
 RA Apelin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;
 RA "Mapping of the human dentin matrix acidic phosphoprotein gene (DMP1) to the dentinogenesis imperfecta type II critical region at chromosome 4q21.";
 RA Genomics 30:347-349(1995).
 CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF EXTRACELLULAR MATRIX.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).


```

[1] SEQUENCE FROM N.A., AND SEQUENCE OF 31-36.
RN RC STRAIN-6715 / UAB66;
RP RC MEDLINE=94292401; PubMed=8021165;
RX RX Wanda S.-Y., Curtiss R. III;
RT RT "purification and characterization of Streptococcus sobrinus
RT RT dextranase produced in recombinant Escherichia coli and sequence
RT RT analysis of the dextranase gene.";
RL RL J. Bacteriol. 176:3839-3850(1994).
CC CC -1- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE
CC CC PELLICLE-COATED TOOTH SURFACE. THE ACTIVITY OF THIS ENZYME IS
CC CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic
CC CC linkages in dextran.
CC CC -1- SUBUNIT: HOMODIMER.
CC CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC CC an amide bond (Potential).
CC CC -1- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
CC CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC use by non-profit institutions. There are no restrictions on its
CC CC modified and this statement as long as its content is in no way
CC CC entities requires a license agreement. Usage by and for commercial
CC CC or send an email to license@isb-sib.ch).
CC CC -----

```

EMBL:	M96978;	AA21772.1;	--
InterPro:	IPR001899;	Gram_pos_anchor.	
Pfam:	PF00746;	Gram_pos_anchor; 1.	
TIGRFAMs:	TIGR01167;	LPTXG_anchor; 1.	
PROSITE:	PS50847;	GRAM_POS_ANCHORING; 1.	
Hydrolase:	Glycosidase; Cell wall;	Peptidoglycan-anchor; Repeat;	
Signal:	Plasmid.		
SIGNAL:	1	30	
CHAIN	31	1308	DEXTRANASE.
PROPEP	1309	1337	REMOVED BY SORTASE (POTENTIAL).
SITE	1305	1309	LPTXG SORTING SIGNAL (POTENTIAL).
MOD_RES	1308	1308	AMIDE-LINKED TO CELL WALL (POTENTIAL).
SEQUENCE	1337 AA;	143298 MW;	B494275&B7273E3.

Query Match 3.8%; Score 171.5; DB 1; Length 1337;
Best Local Similarity 19.8%; Pred. NO. 0.18;
Matches 159; Conservative 96; Mismatches 304; Indels 243; Gaps 34;

471 DPKSKFWLPEGYAEFLRAIKELPNLYLTVD--VN-----GEQIYRLKDGNDQVIYN 521

00 00 PI-----SRKVLTTIR-----SGHRANIESAK 101
 1: || ||| :
 522 EIWPGPALPSEMAVKNTWTSRPVLTCKRQDWNKISVYAWEGSENGSKADAECK 581
 103 00

102 RLF-----CTNDKQIVSSGDDVYITNVEQDAETNRQCQF-----TCHY-----GTYYEIM 148
| | : | : | : ||: | : |
582 SLQTDVAULTSASTAAAGGYHMSLAALANQQDDETGQGIGVLQTAYYTTSLEKWSSELT 641
|||

198 642 RKNNDYQOFITAYENVLRDGVENDQAQVTFDS-----NQKILSTD-- 682

135 TCPPYPLAVGSSDSVRIYDR-----MLGTRATGNAGRTGTGWARETPS 247
683 -----ARGITGQWWTYGGKGNFRTVOLLNMG-----NSDWNKDGSRANKTPD 729

246 HLNKSKRTSLCYG-EDGEILLVSSVDYIYLFDPKDDTARELKTPSAERREELRQP- 305
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
730 EQINLVKYALGVDSMEDAQRM-----ANQIVTSPDWSKSNIQLKVSASVKTDENGPV 784

306

785 LVINVPKRLTL---WDVVYISNANQSAPEAOAQTP---AAQSSDDKVAE--NETSQPAA 836

[illegible]

RESULT 18

ID	YH6_YEAST	STANDARD;	PRT; 1189 AA.
P47035;	01-FEB-1996 (Rel. 33, Created)		
01-FEB-1996 (Rel. 33, Last sequence update)			
15-JUL-1998 (Rel. 36, Last annotation update)			
Hypothetical 128.5 kDa protein in SCPl60-SMC3 intergenic region.			
YJL076W OR J1038.			
Saccharomyces cerevisiae (Baker's yeast).			
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
NCBI_Taxid=4932;			
[1]			
SEQUENCE FROM N.A.			
Rose M., Koetter P., Entian K.D.;			
Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.			
-!- SIMILARITY: TO YEAST YKR010C			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch

EMBL; Z49351; CAA89367.1; SGD; S0003612; YJL076W.
Hypothetical protein.

DOMAIN	192	199	POLY-THR.
DOMAIN	328	335	POLY-SER.
DOMAIN	415	418	POLY-SER.
DOMAIN	982	988	POLY-SER.
DOMAIN	993	1000	POLY-SER.
DOMAIN	1171	1174	POLY-LYS.

```

SEQUENCE      1189 AA; 128530 MW; 80EDF9A8065E5A2 CRC64;
Query Match
Best Local Similarity 3.8%; Score 170.5; DB 1; Length 1189;
Matches 151; Conservative 109; Mismatches 286; Indels 209; Gaps 33;
14 VKRS-----LGLIEDPSRLSRYSYLRGRREFIQRLKLEALINVDGCVN-----TICWNIDTGE 64

```

1

1

[illegible]

Zhang X., Zhao J., Li C., Gao S., Qiu C., Liu P., Wu G., Qiang B.,
Lo W.H.Y., Shen Y.;
"DSP mutation in dentinogenesis imperfecta Shields type II.";
Nat. Genet. 27:151-152(2001).
[4].

VARIANTS DFNA39/DGII1 THR-17 AND PHE-18.
MEDLINE=21096982; PubMed=11715790;
Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,
Zhang J., Shi X., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,
"Dentinogenesis imperfecta 1 with or without progressive hearing loss
is associated with distinct mutations in DSPP.";
Nat. Genet. 27:201-204(2001).
-1- FUNCTION: DSP may be an important factor in dentinogenesis. DPP
may bind high amount of calcium and facilitate initial
mineralization of dentin matrix collagen as well as regulate the
size and shape of the crystals.
-1- SUBCELLULAR LOCATION: Secreted.
-1- TISSUE SPECIFICITY: Expressed in teeth. DPP is synthesized by
odontoblast and transiently expressed by pre-ameloblasts.
-1- PTM: DSP is glycosylated.
-1- DISEASE: Defects in DSPP are the cause of dentinogenesis
imperfecta 1 (DGII1), also known as dentinogenesis imperfecta
Shields type II. It is an autosomal dominant disorder in which
both the primary and the permanent teeth are affected. It occurs
with an incidence of 1:8000 live births. The teeth are amber and
opalescent, the pulp chamber being obliterated by abnormal dentin.
The enamel, although unaffected, tends to fracture, which makes
dentin undergo rapid attrition, leading to shortening of the
teeth.

-1- DISEASE: Defects in DSPP are the cause of autosomal dominant
deafness with dentinogenesis imperfecta 1 syndrome (DFNA39/DGII1
syndrome). Affected individuals present with DGII1 associated with
early onset progressive sensorineural high-frequency hearing loss.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF163151; AAF42472.1; -;
ENFEMBL: AF094508; AAD16120.1; -;
Genew: HGNC:3054; DSPP.
MIM: 125485; -;
MIM: 125490; -;
MIM: 605594; -;
Signal: Extracellular matrix; Glycoprotein; Sialic acid;
Phosphorylation; Deafness; Disease mutation.
DISEASE
1 15
HAIN 16 1253 DENTIN STALOPHOSPHOPROTEIN.
HAIN 16 462 DENTIN STALOPROTEIN.
HAIN 463 1253 DENTIN STALOPROTEIN.
OMAIN 439 1253 SER/ASP-RICH.
SITE 488 490 CELL ATTACHMENT SITE (POTENTIAL).
COD_RES 259 259 PHOSPHORYLATION (BY CK1) (POTENTIAL).
ARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 49 49 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 150 150 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 191 191 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 222 222 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 275 275 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
ARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
VARIANT 17 17 P -> T (IN DFNA39/DGII1).
-1-
18 18 /FTID-VAR.012280.
V -> F (IN DFNA39/DGII1).

QY 511 SHCKSEGO-----EESFVPOSSVOPPEGD-----SETKAPESSSEDVTKYQGVSAENPVEN 562
 Db 1258 KEFVTEKNPVVEE---PSKQEDKPKVEKIVAEQAPVEPEDEEDSTSTTTTET 1314
 QY 563 HINIQTOSDKFTAKP-----LOSNGERNDLNRSCGVPESASS-----EKAKE 607
 Db 1315 P-----SPSGSYTKFGLSGRSDVDEQSGSNKLDYDDNNWQILANRGVDTDLTKSHE 1371
 QY 608 PETSDOTSTESATNEN-----NTNPEPOFO-----TEATGP-----SAHETSRTDRA 650
 Db 1372 RQFAGQONLEPKNCNDGNSQEQYGPYEFQADEPSTTEGKAKRALSVKQQAOLNAA 1431
 QY 651 LQDT-DDSDDD-----DVLTPGARYRAGPGDRSAVARI-----QEFERKRKKEKEME 697
 Db 1432 LNDGSDSSDGTITTTTPSYAMRGKRVRRSTARRVPIPKIGKASDEVWVRSPROAKMP 1491
 QY 698 ELDTLNLRRP 707
 Db 1492 QRPKKSWSRP 1501

RESULT 22
 ID IE63_MCMVS STANDARD; PRT; 841 AA.
 AC Q69154;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Transcriptional regulator IE63 homolog (Protein UL69).
 GN UL69.
 OS Murine cytomegalovirus (strain Smith).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Muromegalovirus.
 OX NCBI_TaxID=10367;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96082764; PubMed=7483291;
 RA Messerle M., Rapp M., Lucin P., Koszinski U.H.;
 RT "Characterization of a conserved gene block in the murine
 RT cytomegalovirus genome."
 RL Virus Genes 10:73-80(1995).
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,
 CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HVS-1 57, HCMV AND MCMV
 CC UL69.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; L07319; AAA96666.1;
 CC Transcription regulation.
 KW
 SQ SEQUENCE 841 AA; 92706 MW; 9F18372739D723B4 CRC64;

Query Match
 Best Local Similarity 3.7%; Score 168; DB 1; Length 841;
 Matches 132; Conservative 61; Mismatches 205; Indels 164; Gaps 30;

QY 190 CRRATSVAIC-----PPI-----PYLAVGSSSVRIYD-----RRML----- 224
 Db 326 CRKGGSSAAAASFDVLRPPFLSKFPLQALALIASFARIVGVRRLSRHSGPFFIND 385
 QY 225 --GTRATGNAGRTGCMVARFTPSH-LNNKSCRVTSLCYSEGGQELIVSYSSDYILFD 281
 Db 386 FDDTGADTSRCGMSELIFDYLPGHRCNEICRVKL-----KLLQPYTS--TLFFC 436
 QY 282 PKDQTARE---LKTPSAEERRELOPPVKRLRLRGLDSDSGTPRARP-----ESER- 329
 Db 437 AYANTRKHPNGLPARSPERRAPDTPNIPRLAYRRS-ATTSPEVEPAPPSPRMTSSSPRV 495

QY 330 -----ERDQSGSPNVLMOR-----MSDMLSRWFEEASEVAOSNRG--RG 367
 Db 496 DSRGGGDRRGDSSSTSNHHRHTRARTRSTHDSGSSRRSSATDGRRRGSRRG 555
 QY 368 RSRPGRGTQSODISTLTPVPSSPDLEYSETAMEVDTPAEQFLOPSTSSWTMSAQAHSTSSP 427
 Db 556 EAQRESNGHHSKS-----PST-----VSSITVHGQCARGDSAPSRKSSQSQOQPETTSK 606
 QY 428 TES-----PHSTPLSSPDSEOROSVEASGHHTHQSDNNNEKLSPK---PGTGPV-L 477
 Db 607 ESSKTAAMPPPPGCPSPSPASRRR-----PSKSPSSSPRPHDPSPSEPADA 653
 QY 478 SLHYSTEG-----TTTSTIKLNTDEWSSSTASSRGISGCKSEGOERSFVPS 526
 Db 654 EKELATAGDEDEGVSPGECVATRRGSADESSDSSSSSSSSSS-----SSDEEEDVEDC 710
 QY 527 SVOPPEGSETKAPESSSEDVTKYQGVSAENPVENHINITQSKDKTAKPLDSNSGRND 586
 Db 711 R-----ELDLQSKRLEALEE-----RCERDFE-----ADDEFAEPI-----EEDD 747
 QY 587 LNLDRSCGVPEESASSEKAKEPETSDQTSTESATNENNTNPEQFQTEATGSA-----HE 642
 Db 748 LH-----CSLMEDEDEPLDPET-ESVWTASVT-----PLAAPPSSIRILDHE 790
 QY 643 ETSTRDSALQDTD--DSDDDPV 662
 Db 791 PGDAEEESDSTDYDQPL 812

RESULT 23
 ID DNJM_MYCPN STANDARD; PRT; 910 AA.
 AC P75354;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE DnaJ-like protein MG200 homolog.
 GN MPN119 OR MP035.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RA MEDLINE=97105885; PubMed=8948633;
 RA Himmelfreid R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae."
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AE000004; AAB95683.1;
 CC HSP; P25685; 1HDJ.
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ; 1.
 DR SMART; SM00271; DnaJ; 1.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS0076; DnaJ_2; 1.
 KW Hypothetical protein; Chapterone; Complete proteome.
 FT DOMAIN 4 73 J-DOMAIN.
 SQ SEQUENCE 910 AA; 100190 MW; 125D0E37D2D221A7 CRC64;

Query Match
 3.7%; Score 168; DB 1; Length 910;

J. Neurochem. 67:89-97(1996).

-|- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
GENE EXPRESSION BY AFFECTING CHROMATIN.

-|- SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEXIN V IN A CALCIUM AND
PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER.

-|- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
INTERACTING WITH HP1 (BY SIMILARITY).

-|- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; D64059; BAA10936.1; --
KW DNA repair; Nuclear protein; DNA-binding; Helicase.
FT NON_TER 1 POLY-SRR.
FT DOMAIN 251 256 POLY-SRR.
FT DOMAIN 304 307 POLY-SRR.
FT DOMAIN 360 367 POLY-ASP.
FT NON_TER 527 527
FT SEQUENCE 527 AA; 59258 MW; ABFA4B10C086D638 CRC64;

Query Match 3.7%; Score 167.5; DB 1; Length 527;
Best Local Similarity 20.8%; Pred. No. 0.096; Indels 121; Gaps 20;
Matches 108; Conservative 83; Mismatches 206;

QY 221 RMLGTRATGNAGRTGTGMVARFIPSHLNKNSCRVTSLCYSDESDGEILVSYSDDIYLF 280
Db 86 KQRTGR-----GKKSTDLKEEKVKREYES-----SSDGTEKL--EGEEIGLF 129
QY 281 DP-----KDDTARELK-----TPSAERRRELPOPPVKRLRLRGWSDTGPRARPESERE 330
Db 130 SKGVQKNKDITDEAKKGKKWKDKCKEKBEL-SDSVDRLPVKGESCDS-----EDKKT 183
QY 331 RDGEOSPYNVLQMRSMDLNRWFPEASVAQSNRGRSRPRGGTQSODISTLTPTVFSP 390
Db 184 RN-----RVSLAEK-----KQFLPAKSS-----GKRPECSSTDTERSVKGECCDST 225
QY 391 DLEVSETAMEVDTPAQFLPQTSTMTGAQAHSSTPSHPSTPLLSPPDSEQRQSVEA 450
Db 226 DKVRKIDL-----ERRSNSKRSTKVYSGSSSSDAEG-----SSDAKKOKKORM 273
QY 451 SGHHTHQSDNNNEKLSPPGPGTEPVLSLHYSTEGTTTKTLNFTDEWSSTASSRGIG 510
Db 274 SA-----KKNSNTKERKK-----SLRATTTKRKQADITSSSDICDDONSA 317
QY 511 SHCKSEQEISFPQSSVOQPE-GDSETKAPERSSDVTKYQEGVSAENPVNHINITQS 569
Db 318 GEESDEQIKPVTLENLVLSHTGFCQSGDEAFSKVPATVDDDDNDNDPENRI----A 373
QY 570 DKFTAPLDSNGERNDLNLRSCGVPEESASSEKAKEPETSDOTSTESATINENNT---- 625
Db 374 KXMLEEIKAN-----LSSDEDGSSDDEPEKEGKKRIGKQSEETPGDCSQNVNSESDS 427
QY 626 ----NPEQFQ-----TEATGPSAHETSTRSALQTDSDDDPVLII 664
Db 428 DSESKKPYRHLRLHKLSLDGSESGEKKTKPKHEKTKGNRRKVSSESDSDTFQE 487
QY 665 PGARYAG-----PGDRRSAVARIQEFFRRKERKE 695
Db 488 SGVSEVSESEDQRPRTRSARKKALEENQSYKKKK 525

RESULT 25		
NNP14_RAT	STANDARD;	PRT; 704 AA.
ID NP14_RAT		
AC P4177;		
DT 01-NOV-1995	(Rel. 32, Created)	

	Best Local Similarity	20.5%;	Pred: NO. 0.17;	Gaps	32;
	Matches	139;	Conservative	Mismatches	89;
				Indels	186;
QY	261	YSDGQELLYVSYDSDIYL-----FDPKDDTARELKTPSAERREELRQ-----PPVKRL	310		
Dd	124	WKSQKEESTSTTEYADVADGLIEDYPPOSDYDDI--PDVDARIEEDVSAYADDIPDV	181		
QY	311	RURGOWSGTGPRAPESERERD-----GEQSPNVSLMQRK-----SDMLSR	351		
Dd	182	DAGMDWQNAEVANSASELIPDVDAGLADEFNTSSAQPASQAWMEAMIGNPEYGYFDAAGE	241		
QY	352	W-----FEBA-----SEVAQSNRGSRPRGGTQSODISTLTPTVPSPDLVESETA--	398		
Dd	242	NNWKGFDEAGQWWLEETEPSSYSNDET-----TTSDAVTAATTAATEEEDQQ-SWANS	295		
QY	399	---MEVDTPAEQFLQPTSS-----TMAQAHSVSSPTEPHSPLSLSSPSQRSO----	447		
Dd	296	APEVDVETPVE--LOPETEPEPIILS-----SEPVEAPASVVIEPTPEIBETTSAVE	347		
QY	448	VEAS-GHHTHHOSDNNEKLSPKPGTGEVLVSHYSTEGTITTSIKLNFTDEWSSTASS	506		
Dd	348	MDASVKADVDEADATNE-----PTQDTISEPEQTDAALAEINHHTADLEPAEVSAT	402		
QY	507	RGISHCKSEQEESFYQSQVOPPEGDSKETKAPESEDVTKYQG-----VS	555		
Dd	403	NDL-----EQDVVEKFNSEPESTDVTDATDPVWEQATETSTNGKFFNFSSFVLS-	453		
QY	556	AENPVENHINITQSKDTAKP--LDNSCG-----RNDNLNDRSCGV-----	595		
Dd	454	DQNP-NPQPTPHHEEDAAPPEIVDTSGESTAPEVTIAESTVELETAAEINNPAITVE	511		
QY	596	---PEESASSKAKEPETSQDTSTESATNNNTNPFOQTGATGPSAHEETSTRDAL	651		
Dd	512	EYLPQTKTVVDKIDPTVAKPTVSDS-----ENSVAPEPEF---VAGF---EQTFSMKP	652		
QY	652	QTDGD-----SDDDPVLIFGARYRAGDGRRRSNAVARIQEFFERRRKERRK	694		
Dd	563	SETEEEPLTAVEPASETOQLIAEDVTSPVTPATAIPAPSINAVPTPAVETFAEAVD--	620		
QY	695	EMEELDNLTRIRPLVMKVYKGHNSTRMIKEANFWGANFMVMSDCGHIFWD-----R	748		
Dd	621	FLKEAAKIEAQPLVPTVPE-----QIDGTDPSSLTOQWDEYLEKTR	661		
QY	749	HTAEHLMLEADNHVACLOHPHFDPLLASSGTDYDIKIWSP-----LEESRIFNRK	800		
Dd	662	KLFHKLFLEQPLPTVKTQDFEIVDPNLDHENVNLIYTEHVQICFLNQLEKEIR-YTRK	720		
QY	801	LAD---EYITRNELMLE	814		
Dd	721	LVDPQTQVTTTESITLE	737		
	RESULT 24				
	ID	ATTR_RAT	STANDARD:	PRT:	527 AA.
AC	P70486:				
DT	15-JUN-2002 (Rel. 41, Created)				
DT	15-JUN-2002 (Rel. 41, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
					(pABP-2)

NCBI_TaxID: 60680,
[1] SEQUENCE FROM N. A., AND CHARACTERIZATION.
STRAIN-Wistar; TISSUE-Embryonic brain;
MEDLINE-96301899; PubMed-8667030;
Ohsawa K, Imai Y, Ito D, Kohsaka S;
"Molecular cloning and characterization of annexin v-binding proteins
with highly hydrophilic peptide structure.";

[illegible]

Query Match

[illegible]

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE Chondroitin sulfate proteoglycan core protein 2 (PG-M) (Glial
 GN hyaluronate-binding protein) (CHAP).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105186; PubMed=7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RX TISSUE=Placenta;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RN EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RX TISSUE=Glial tumor;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 RL alternative splice variants of human versican.";
 RN J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RX TISSUE=Lung fibroblast;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RL lectin-like and growth factor-like sequences.";
 RN J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RL human chromosome 5 (5q12-5q14).";
 RN Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RX TISSUE=Brain;
 RA MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RL without a chondroitin sulfate attachment in region in mouse and human
 RN tissues.";
 RN J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RX TISSUE=Aortic smooth muscle;
 RA MEDLINE=99327053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wright T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RX TISSUE=Brain;
 RA MEDLINE=89174663; PubMed=2466833;
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 RL hyaluronate-binding protein.";
 RN J. Biol. Chem. 264:5981-5987(1989).

RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -|- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -|- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -|- ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,
 CC V2, V3 and Vint; are produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -|- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -|- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -|- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -|- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U16306; AAA65018.1; -
 CC EMBL; X15998; CAA34128.1; -
 CC EMBL; S52488; AAB24878.1; -
 CC EMBL; U26555; AAA67565.1; -
 CC EMBL; D32039; BAA06801.1; -
 CC EMBL; J02814; AAA36437.1; -
 CC EMBL; AF084545; AAD48545.1; -
 CC PIR; S06014; S06014. -
 CC PIR; A29348; A29348. -
 CC PIR; A30358; A30358. -
 CC HSP; P01132; IEGF. -
 CC Genew; HGNC:2464; CSPG2. -
 CC MIM; 118661; -
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR001438; EGF_II.
 CC InterPro; IPR003599; Ig_II.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; Lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC PRINTS; PR00010; EGFBLD.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00001; EGF_like; 1.
 CC SMART; SM00409; IG_1; 1.
 CC SMART; SM00445; LINK; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 1.

Query Match	3.7%;	Score 166.5;	DB 1;	Length 3396;
Best Local Similarity	20.8%;	Pred. No. 0.94;		
Matches 145;	Conservative 72;	Mismatches 240;	Indels 239;	Gaps 28;
138	TCHYGTYTEIWPNDPYTFLSCGEGDGTWVFD	---TRIKTCTKEDKDDJILNCRRAA	194	
545	TGHYGF---	LGEBDDSDRTLLVGSDESLTFIQIPEVTVKTS	EDTITHLEDLESVS	601
195	TSVAICPPIPYLLAVGSDSVRIYDRMLGTRATGNAGR	---GTTGMVARIPISHLNKK	252	
602	ASTTVSP----	LIMPNDNGSSMDWEERQTSGRITFEFLGYLSTT	---PFFSQH----	649
253	SCRVTSLCYSDGEILVSVSSDYIYLFDPKDJTAR	-----ELKTPS-AEEREEL	302	
650	--RTEIELFPYSGKILVEGISTVIYPSLOTENTHRRERTETILPEMRDTYDEIOEII	707		
303	RQPVKRLRLGDSWDTGPRAPESERERCEQSPNYSLMORMSDMLSRWFSEASEVAQS	362		
708	TASPPM-----	---GKTEEEVFSGMKLSLSTLS----	EPIHVTES	739
				414

Db 731 --LHSDYKNCMTM--CP--EVKDPISIRNNEQTSQAVPEENTEIAHRTETPSISEGP 784
QY 432 HSTPLLSFDSQROSVASGHHHQQSDNNNEKLSKPKGPGTGEVLSLHYSTEGTTSTI 491
Db 785 -----KGNQEKERDDSLSKTSVSPENSRP-----ETDAKDTISNL 819
QY 492 -----KLANFTDEWSSIASRRGIGSHCKSEGOEFSFVPOSSVQ 529
Db 820 LKLKGDVDSMPKAVIGSPDNINVTGQDNIS-----GVNSKPLSD-----DVR 865
QY 530 PPEGSETKAPESSEDVTKYQGVSAENPVENHINIQSDKFTAKPLDSNSG-----ER 584
Db 866 PDKKELEDQNSDESETVYVNH-----ISKSPSINNDDSGSGSATVSESSSNTGLSIDDDR 922
QY 585 NDNLDRSGVGPESASSEKAKEPESDQTSSTES-----ATNENNTNPEPQFQTEATG 637
Db 923 NGDTFFVKT-----QDTANTEDVIRKENADKDEDEKGADEHERHSTSELSPEEKMLTDNEG 978
QY 638 PSA--HEE-----TSTRDSA-----LQDT-----DD-----656
Db 979 GNSLNHEEVKEHTNSDNVQSGGIYVMNVKELKDTLENPSSLDGKAHELSEPNLS 1038
QY 657 SDDPVLIPGARYAGGDRSARVATQEPFRR--RKERKEMEELDTLIRPLVKNMYK 714
Db 1039 SQQDMSNTP-----GPLDNTS-----EETTERISNNEYKVNREDEBTLTKKEYEDIVLK 1087
QY 715 GHRNSR-----TMKEANFWANFVMSGDCGHFIWDRHTAHLMLLEAD 760
Db 1088 SHMNRSDGELYDENSGLSTVNDSEF--DAEAKMGNDTSEM--SHNSSQH--IESD 1139
QY 761 -----NHVNCLOPHPPDPILASSGIDYDI-----KIWSPLESRIFN-- 798
Db 1140 QOKNDMTVGLGTHVQNEIS-----VPVTGEIDKLESKESKI-HKAEERLSHTD 1192
QY 799 --RKLADDEVITRNELMLETRN 818
Db 1193 IHKINPEDRNSNTLHLKDNRN 1213
RESULT 31
ID ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain stem;
RC MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes."
RL J. Cell Biol. 114:241-253(1991).
RN [2]
RP REVISIONS.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons."
RL J. Cell Biol. 123:1463-1473(1993).

[4]
RN SEQUENCE OF 463-495 FROM N.A.
RP MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Peng T.L., Francke U., Sahr K.E.,
RT Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene."
RL Genomics 10:858-866(1991).
CC -1- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC elements. Also bind to cytoskeletal proteins.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC CELLS THROUGHOUT THE BRAIN.
CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC AND FUNCTION (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56957; CAA40278.1; -
CC EMBL; X56958; CAA40279.2; -
CC EMBL; Z26634; CAB42644.1; -
CC EMBL; M37123; AAA62828.1; -
CC PIR; S14533; S14533.
CC PIR; A39643; A39643.
CC PIR; B39643; B39643.
CC PIR; S14569; S14569.
CC HSP; P42771; IDC2.
CC Genew; HGNC:493; ANK2.
CC MIM; 106410; -
CC InterPro; IPR002110; ANK.
CC InterPro; IPR000488; Death.
CC InterPro; IPR000906; ZUS.
CC Pfam; PF00023; ank; 24.
CC Pfam; PF00531; death; 1.
CC Pfam; PF00791; ZUS; 1.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 21.
CC SMART; SM00005; DEATH; 1.
CC SMART; SM00218; ZUS; 1.
CC PROSITE; PS50088; ANK_REPEAT; 20.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation.
KW REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.

FT	REPEAT	694	723	ANK 20.
FT	REPEAT	727	756	ANK 21.
FT	REPEAT	760	789	ANK 22.
FT	REPEAT	793	822	ANK 23.
FT	DOMAIN	1773	1950	REPEAT-RICH REGION.
FT	REPEAT	1773	1784	REPEAT A.
FT	REPEAT	1785	1796	REPEAT A.
FT	REPEAT	1797	1808	REPEAT A.
FT	REPEAT	1809	1820	REPEAT A.
FT	REPEAT	1821	1832	REPEAT A.
FT	REPEAT	1833	1844	REPEAT A.
FT	REPEAT	1845	1856	REPEAT A.
FT	REPEAT	1857	1867	REPEAT A. (APPROXIMATE).
FT	REPEAT	1868	1879	REPEAT A.
FT	REPEAT	1880	1891	REPEAT A.
FT	REPEAT	1892	1902	REPEAT A.
FT	REPEAT	1903	1914	REPEAT A. (APPROXIMATE).
FT	REPEAT	1915	1926	REPEAT A.
FT	REPEAT	1927	1938	REPEAT A.
FT	REPEAT	1939	1950	REPEAT A.
FT	DOMAIN	3536	3620	DEATH.
FT	VARSPLIC	1039	1039	
FT	VARSPLIC	1444	3528	
FT	CONFLICT	475	476	
FT	CONFLICT	971	971	
FT	CONFLICT	3581	3582	
FT	CONFLICT	3586	3586	
SQ	SEQUENCE	3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;		

Query Match

Best local Similarity 3.7%; Score 165.5; DB 1; Length 3924;

Matches 111; Conservative 100; Mismatches 265; Indels 127; Gaps 22;

QY	293	TPSAERREELRQPPVKRLRLRGDWSGTGPPRAPERERDGEQ-----SPNYSLMQRM 345
DB	2507	TPSSEVEYVTPKTT-----DVSTPKPAVIECAEEDDSENGEKKRFTPEEMFKW 2559
QY	346	SDMLSRRWFEASEVAQSNRGR-----SRRPGGTQSODISTLP 384
DB	2560	TKI--KMFDELEAKOKRDYKKEPKQEISSSSDDPDADCSVDYDEKFKHGSGEDESGVP 2617
QY	385	TVPSPDLEVSETAMEVDTPAQFLQSTSTWSAQAHSTSPTESPHSTPLSSPDSFSEQ 444
DB	2618	VLVTSERKVSSSS-ESEPELAQLKKGADEGLLPPEYIRVQPPSLPSSMDNSNPEEQ 2676
QY	445	RQSYEAGHHTHQSDNNNEKLSPKPGTGPVLSLH-----YSTEGTTTSTIKLNF 495
DB	2677	FQPV-VSKQVTFKNNDTQE-----EPCKSEEKDSHSLAEDHVAVSTEADRSYDKLNR 2731
QY	496	TDEWSSITASSRGIHGCKSEGEQEEFVPOSSVOPPEGDSSETKAPESSESDVTKYQEGVS 555
DB	2732	DTDQPKICD-----GHGC-----EAMSPSSARPVSSGLQSPGTGDVDQEPVYIKSLA 2780
QY	556	AENPVENHINITQDKFTAKPLDNGSERNDLDRSCGVPESASSEKAKEPESDQTS 615
DB	2781	LOG---THEKDTGEELDVSAESPQADCPSEFSFSSSLPHCLVSEKLEDEDSATS 2837
QY	616	TE---SATNENNTPEPQFQFATQPSAHETSTRDALQTDSD--DPVLIPGARY 669
DB	2838	IQTEVTKTIDETENLKPDCPSODSITQTDRFSMDVPVSDLAENDEIYDQIIT--SPY 2895
QY	670	RAGPGDRRSAVARIQIEFFRRRKRKEME-----ELDTLNIIRPLKVMYVYKGRHS 719
DB	2896	ENVPs-----QSFSSSESKTQTDANHTTSHSSESVISVITSPVEDVYVASSSG 2946
QY	720	RTMIKEANFWGANFYMSGSDGCHIFITWRHRTAEHLMLLEADNHVNVNCLQPHFPDIPILAS- 778
DB	2947	TVLKSNFEGQDIK-----ESQLESTLWQSDS-----VSSSEFETMSAT 2989
QY	779	-----SGIDYDIKWSPLEESRIFNKLADVITRNEMLME-ETRNIT--T 821
DB	2990	TTVVGQISKVLIITKDVSWSIREDEAEFARVKEEQKIFGLMVDQKSOQTTPDT 3049


```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49198; CAA89059.1; -
CC EMBL; U20618; AAB64526.1; -
CC SGD; S0004322; CHS5.
CC InterPro; IPR00137; BRCT.
CC InterPro; IPR003961; FN_III.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00533; BRCT; 1.
CC SMART; SM00292; BRCT; 1.
CC SMART; SM00060; FN3; 1.
CC PROSITE; PS00172; BRCT; UNKNOWN_1.
CC Repeat.
CC KW DOMAIN 76 166 FIBRONECTIN TYPE-III.
CC FT SEQUENCE 671 AA; 73638 MW; FA92741B862814C2 CRC64;
CC SQ

```

Query Match 3.6%; Score 161.5; DB 1; Length 671;
Best Local Similarity 20.9%; Pred. No. 0.25;
Matches 84; Conservative 59; Mismatches 141; Indels 117; Gaps 14;

```

QY 283 KDTARELKTSAEERREELRQPPVKRLRLGRDWSDTGPRARPESEERDGEQSPNVSILM 342
DB 351 ENETITVNPVS---RSLKSEPVGTNIEENKADSAEAVVEEPNEVAESSNEEAT 406
QY 343 QRMS-----DMLSRWFEEASEVAQSN---RGRGRPRG-----GTSQSDISTLP 384
DB 407 GKSEDTTHNEQADNGFVQTEVAENNIITESAGENNEPADDAAAMEFGREAEIET-- 464
QY 385 TVPSSFDLEVSETAMEVDTAPQFQLPQSTSTMSQAQHSSTPSTPSTPLLSPDSEQ 444
DB 465 -----PEVNESIEDANEPAEDSNPEVDSNKPVK---DSNKPVEDS-----NKPVEDS 509
QY 445 QSVFASGHHTHQSDNNNEKLSRPGTGPVLSHYSTEGTTTSTIKLNFIDEWSSTAS 504
DB 510 NKPVEDSNKPVEDANEPEVDETSEVEDAGEPVQE-----TNEFTTIDIA 552
QY 505 SSRGIGSHCKSEGESEFVQSPVQPPGSDSETKAPESSESDVTKYQGVSAENPVENHI 564
DB 553 SPR-----HQEDIELEAE---PKDATESVAVPSNEDVKPEEKGEAEADDINN-- 598
QY 565 NITQSKDKTAKPLDSNGERNDNLDRSCGVPEESASSEKAKEPETSDOTSTESATNENN 624
DB 599 -----VSKEASGE-----STHQKTEASASLES 623
QY 625 TNPEQFQTEATGPSAHEETSTRD-----SALQDTDDSDDD 660
DB 624 AVTEEQETEA-----EVTDDVLSTKEAKKNTGNSNSN 657

```

RESULT 36
HET1_PODAN STANDARD; PRT; 1356 AA.
ID HET1_PODAN
AC Q00808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetatible incompatibility protein HET-E-1.
GN HET-E1.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96009891; PubMed=7557402;
RA Saupe S., Turcq B., Begueret J.;

FT REPEAT 465 494 ANK 13.
FT REPEAT 498 527 ANK 14.
FT REPEAT 531 560 ANK 15.
FT REPEAT 564 593 ANK 16.
FT REPEAT 597 626 ANK 17.
FT REPEAT 630 659 ANK 18.
FT REPEAT 663 692 ANK 19.
FT REPEAT 696 725 ANK 20.
FT REPEAT 729 758 ANK 21.
FT REPEAT 762 791 ANK 22.
FT REPEAT 795 825 ANK 23.
FT DOMAIN 1519 1898 SER-RICH.
FT DOMAIN 4090 4174 DEATH.
SQ SEQUENCE 4377 AA: 480399 MW: F42379E55768B684 CRC64;

Query Match 3.5%; Score 160.5; DB 1; Length 4377;
Best Local Similarity 20.6%; Pred. No. 2.5;
Matches 192; Conservative 100; Mismatches 327; Indels 315; Gaps 47;

QY 55 NTICWMDTGEVILSGSDTKLVISNPYSRVKLTITRSHRANIFS---AKFLPC--TNDK 109
DB 2797 NEIIVND-----SGSDNVK-----KQTEMSSKAMPDFSEQQAKLACHITS- 2840
QY 110 QIVSCSG--DGVIFYNVEQDAETNRQC-----FTCHYG-----TYEI 147
DB 2841 --LATRGFWKVKFRT--WESSGATNNKSQEKLSHLVLDVRENHIGHPESKSDQKNEF 2897
QY 148 MTPVNDPYTFLSCGEGDTVRFDFRIKTSCTK-----EDCKDILINCRAATSV 197
DB 2898 MSVTEREKLLTNGLSLEIK--EMTVASPKSKVLYREYVVKEGDHPGGLDQPSRSESS 2955
QY 198 AICPPPIYLAVGSDSVRIYD--RRMLGTRATNGYAGRTGTGMVARFIPSH--LNKK--- 252
DB 2956 AV-----SHIPVRVADERRMLSSNIPDGFCEQSA-----FPKHLSQKLSQ 2996
QY 253 -----SCRVTSLCYSEDQEILVYSDDIYVLFDPKDTARELKTSPSAEER 298
DB 2997 SSMSETVETQHFNSIEDEKVTYSEIK--VSKHQSTVGLCPPLEET----- 3041
QY 299 REELRQPVKRLRLGWDSTGPRARPERERDGEOSPNSVLMQKSDMLSR-----WFE 354
DB 3042 -----ETSPKSPDLSFSPGKESPS-----SDVFDHSPIDGLE 3075
QY 355 EASEVAQSNRGR-----CRSPRGG-----TSQSDISTL--- 383
DB 3076 KLAPLAQTEGGKEIKTLPVVYFVQGVQKYEIKIQGGGVKKIIQCEKTVQETRGTFYTT 3135
QY 384 -----PIVPSPDLEVSETAMEVDTPAEQFLQFLOPSTSTMSAQAHSTSPTES----- 430
DB 3136 RQOKPPSPQSPEDDTLEQVFLDSSGKSLPTPTPSESVSEYFTSKTDPDLAYIPG 3195
QY 431 -PHSTPLILSSPDSEQRQSVASGHHT-----HHQSDNNN-----EKL 466
DB 3196 KPSPIPEVSESEEEQAKSTSLKQTTVEATAVERMPNDVSKDSNQRPKNRVAYIEFP 3255
QY 467 SPKPGTGPVLVLS---LHVTGTTTSTTKLNTFDWSSSTASSRIGSHCKSEGOEESFV 523
DB 3256 PPPPLDADQIESDKKHHLPE-KEYDMIEVNLQDE-----HDKYQLAE--- 3297
QY 524 POSSVQPPGEGDSETRAPRESSSDVTKYQBGVSAENPVENHINITQSKFTAKPLDSNSGE 583
DB 3298 PVIRVQPPSPVPPGADVSDSSDESIQ-----PVPVKY-----TFKLEVDDEQKE 3345
QY 584 RNDLNLDRSCGVPEPSASEKA---KEPPTS-----DOTSTESATNENNTNPE 628
DB 3346 K-----PKASAEKASQKESNGSKDNEFGLGLDSPQNEIAQNGNNDQSI 3392
QY 629 PQFQTEATGPSAHEETSTR--DSA-----LQDQD---SDDDPVLIFGARVRAG----- 672
DB 3393 TFCSTATTAEFSDHTDAIDSIDGLDGLDGLTESKULPIQAMETKDKIWNTEGIL 3452
QY 673 -PGDRSARVATQEFRRRKRERKEMEELDTLNRPLVKNVYKGRNRSRTMIKEANFWGA 731

DB 3453 KPADRSFSQSKLEVI---EEEGKVGPD-----KPPSKSSSEKTPDKDQSGA 3500
QY 732 N-FVMGSGDCGHIFIMDRHTAEHMLLEADNHYVNCVLPHPDFPILASSGIDYDIKWS 790
DB 3501 OPFTLSE-----RHPDRSVFDPDTVFSYKVDSEFATPFKTV-ATKGLDFD--PWSN 3547
QY 791 LE-ESRIFNKKLADEVITRNELMLEETRTNITVP 823
DB 3548 NRGDDEVFDSKSRDEDETKPFLAVED-RSPATTP 3580

RESULT 38
TGN2_HUMAN
ID TGN2_HUMAN STANDARD; PRT: 480 AA.
AC Q43493; Q43499; Q43500; Q15282; Q92760; Q43492; Q43501;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trans-Golgi network integral membrane protein 2 precursor (Trans-
Golgi network protein TGN51) (TGN46) (TGN48) (TGN38 homolog).
GN TGN46 OR TGN51 OR TGN48.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS OF TYR-430.
RC TISSUE=Liver, and Placenta;
RX MEDLINE=9086273; PubMed=9422759;
RA Kain R., Angata K., Kerjaschki D., Fukuda M.;
RT "Molecular cloning and expression of a novel human trans-Golgi network
glycoprotein, TGN51, that contains multiple tyrosine-containing
motifs.",
RL J. Biol. Chem. 273:981-988(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM TGN46).
RC TISSUE=Fetal liver, and Fetal thymus;
RX MEDLINE=97063845; PubMed=8907352;
RA Ponnambalam S., Girotti M., Vaspo M.-L., Owen C.E., Perry A.C.,
Suganuma T., Nilsson T., Fried M., Banting G., Warren G.;
RT "Primate homologues of rat TGN38: primary structure, expression and
functional implications",
RL J. Cell Sci. 109:675-685(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
FROM TRANS-GOLGI NETWORK.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
SURFACE RETURNING VIA ENDOSOMES.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TGN46, TGN48 AND TGN51 (SHOWN
HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ISOFORM TGN46 IS WIDELY EXPRESSED. ISOFORM
TGN51 IS MORE ABUNDANT IN FETAL LUNG AND KIDNEY. ISOFORM TGN48
IS BARELY EXPRESSED IN EMBRYONIC KIDNEY AND PROMYELOCYTIC CELLS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AF027516; AAC39542.1; -
CC EMBL; AF029316; AAB96908.1; -
CC EMBL; AF029313; AAB96908.1; JOINED.
CC EMBL; AF029314; AAB96908.1; JOINED.
CC EMBL; AF029315; AAB96908.1; JOINED.
CC EMBL; AF029316; AAB96906.1; -
CC EMBL; AF029313; AAB96906.1; JOINED.
CC EMBL; AF029314; AAB96906.1; JOINED.
CC EMBL; AF029315; AAB96906.1; JOINED.
CC EMBL; U62390; AAC39539.1; -
CC EMBL; AF027515; AAC39541.1; -

QY 438 SSPDSEQRQVEASGHHTHHQSDDNNK-----LSPKPGTGEPVLHLHYSTEGTTSTIKL 493
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
155 GAEPKTQKDSPSKSGSAQAOTTKVPNKGADGQTTPKDGS-----KS 196

QY 494 NFDWSSSIASSRGIGSHCKSBGOEERFPQSSVQPEGDSETKAPESSD----- 546
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
197 GAEDTOTPKDPVNKSGAEAKQTPKDGSKNGABEQ--PIDGPSKGAEOITKSDSPNKVVP 254

QY 547 --VTKYQEGVAENPVNH-----INITQDKFTAKPLDSDNSGERNDLNLRDCSGVPE 597
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
255 EOPSRKDHKSPINPDGNKELPADNQNLADKGLSHPAFKTESGEETDL----- 304

QY 598 ESASSKAKEPETSODTSTESATNENNTPN----POFOTEATGPSAHEISTEDSLQD 653
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
305 ISPPQBEVKSSPETDVGPKEA-EDDDTGPBGSPKPEEKKGSGASSE--NREGTLSD 361

QY 654 TDDSDDDPVLIPCARYPAGCDRRSAVARIQEFRRRKKERKEMEELDNLIRPLVKMWY 713
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
362 STGSEKDDL-----YPNCSG--NGSAESSHF-----AYLVTAI---LVAVLY 400

QY 714 KGRHSRTMIKEANFGANFVMVG 737
Db : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
401 IAHNKRKIITA-----FVLEG 416

RESULT 39
BUD4_YEAST STANDARD; PRT: 1447 AA.
AC P47136;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bud site selection protein BUD4.
DB BUD4 OR YJR092W OR J1905.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=96295555; PubMed=8707826;
RA Sanders S.L., Herskowitz I.,
RT "The BUD4 protein of yeast, required for axial budding, is localized
to the mother/BUD neck in a cell cycle-dependent manner.";
RL J. Cell Biol. 134:413-427(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Ramezani Rad M., Kirchrath L., Hollenberg C.P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: CO-ASSEMBLES WITH BUD3 AT BUD SITES. BUD4 AND BUD3 MAY
COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC -! SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -! CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL
SECTION DUE TO FRAMESHIFTS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licensese@isb-sib.ch).

EMBL; U41641; AAB17116.1; -
EMBL; Z49592; CAA89620.1; ALT_FRAME.
EMBL; Z49591; CAA89619.1; -
SGD; S0003852; BUD4.
InterPro: IPR001849; PH.
Pfam: PF00169; PH; 1.
SMART; SM00233; PH; 1.
PROSITE; PS50003; PH_DOMAIN; 1.

KW ATP-binding. 181 189 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1175 1182 ATP (POTENTIAL).
FT NP_BIND 1302 1413 PH.
FT DOMAIN 291 328 PRASVEDLNINLPCADSSONNPVTTDADALIENDV ->
FT CONFLICT ANSLCGGFKHNESSRVLPDAKIQSLMRRLKTMLC
FT (IN REF. 2).
FT CONFLICT 340 340 D -> E (IN REF. 2).
FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
Query Match 3.5%; Score 159.5; DB 1; Length 1447;
Best Local Similarity 18.8%; Pred. No. 0.77; Indels 289; Gaps 43;
Matches 162; Conservative 127; Mismatches 284; Indels 289; Gaps 43;
QY 213 DSSVRIYDRMLGTRATGNTGVARGTGQVAFIPSHLNKNSCRVTSICYSDDGQIILVSY 272
DB 47 DDYMEVLKHNTRSNATENSNGRSPSKM-----STISNESLNGLLRVNSELESPA 100
QY 273 SSDVI-----YLFDPK-----DDTARELKTPSAERREELROPPVK-- 308
DB 101 HOERIKNSVANGALGHANSPKVLNKNMAQIDIKLAN-----EKPVKLSPLKFT 154
QY 309 -----RLRLGDSVDTGPRARPERERD-----PQSPNV----- 339
DB 155 LKSTQPLLSYPSPHRSIIETNYDD-----EDEEDAYTCLTQSPQILHSPSRI 207
QY 340 -----SLMQRSDMLSRWFEEASEVAQSNRGRSRPRGGTQSQDITSLP 384
DB 208 PITNAVSIKINLNDITLNPNSDK-----SLVSDTFSVDSTGR-----ELDTKIP 252
QY 385 TVP-----SSPDLVSETAMEVDTPAEQFLQ-PTSTSTMSAQAHSTS-SPTSPHSTPL-- 436
DB 253 ELPPCMSSTPEM-----TPVDEKCNLPKLLNTSNNSHSDSRSPASVEDLNIST 302
QY 437 -LSSPDSEQROSVSEAG-----HHTHQSDNNNEKLSKPKG-TCEPVL----- 477
DB 303 NLPAGDSSONNPVTTDADALIENDVVDLQNMHEIDAFDEKVLDECSNEPVTFLGE 362
QY 478 -----SLHVSFTEGTTSTTK-----LNFT--DEWSSIASSSRGIGSHCKS 515
DB 363 NDTIRSVSNKTNANVQEFSDSLAHSEPKFKNLNTSDVWNEKEDKTDANISTSKS 422
QY 516 EQQESFV-----PQSSVQPPGDSGSETKAPEESED----- 546
DB 423 ---EESYIADYKVTRQEDWDTKKLHOESEHANEQPAIIPKQDSSEFTFELANNESEFQ 479
QY 547 -----VTKYQE--GVSAPNPNVHNITQ--SDKFTA-----KPLDSNSGERNDLNL 589
DB 480 FKDGEEYRIVQHEESLYGQRTKSPENIINGSEIGYDGHGAEEVNEPLAKTSAEHDLS- 538
QY 590 DRSCGVPEESASSE-----KAKEPETSQDT-----STESATNNTNPEPQF----- 631
DB 539 -SSC---EQQSVSEARNKRIEKEVETKDNITETERDESEYHKVEENEEPERHVPPLPL 594
QY 632 ---QTEATGPSAHEETSFRDSA-----LQDTHDSDDDVPLIPGARYA 671
DB 595 PRWEETQFNEPFDENDTSDSDTLFRSMKPSDYISLWHLQEEIKNSPESSTANSQF-- 652
QY 672 GPDDRSAVARIQEFFRRKEREKEMELDTLNIRRLPVKMYKGHNSRRTMIKEANFW-G 730
DB 653 ---SQQSSITTASTVDSKDKNGSTSFKEKPRIVSR---SRIY---NPKSRVSSLNATYDN 702
QY 731 ANFVMSGDCGHFIWDRHTAHLMLLEADNHV-----VNCLOPHFPDFPLASSGID-- 782
DB 703 EDYILSNSEWNAIDPMRRNT---LISKRIQDNITQKHAPLIRP-----SIMKINGEDSG 755
QY 783 ---YDIKWSPLEESRI-FNRKLADDEVITRNELMLEETRNITIV-PASFMLRMLASLNHI 837
DB 756 FQNHFLVEQPOEHEHNPPLTHLSEQDITNVGLDEQKLPTNQDEAEISIRETESAGDI 815
QY 838 ---RADRLEGDRSRGSGQENEN 856
DB 816 TFNRGDLISLDFDELQGFAN 837

RESULT 40
YAU9_SCHPO
ID YAU9_SCHPO STANDARD; PRT; 1275 AA.
AC Q10164;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C26A3.09c in chromosome I.
GN SPAC26A3.09c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
[J]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gantley S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Oliver K., O'Neill S., Saunders D., Seeger K., Sharp S.,
RA Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,
RA Skelton J., Simmonds M., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Grymonprez B.,
RA Woodward J., Voicakert G., Aert R., Robben J., Schaefer M., Mueller-Auer S.,
RA Weltjens I., Vanstraelen E., Rieger M., Holzer E., Moestl D., Hilbert H.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Leirach H., Reinhardt R., Pohl T.M.,
RA Borzym K., Langer I., Beck A., Lehmann W., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet W., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z69240; CAA93232.1; -.
CC HSP: Q07960; 1RGP.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000198; RHOAP.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00620; RHOAP; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00324; RHOAP; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Hypothetical protein.
FT DOMAIN 719 836 PH.
SQ SEQUENCE 1275 AA; 143586 MW; 136DA1A26B4A1BBE CRC64;
Query Match 3.5%; Score 159; DB 1; Length 1275;
Best Local Similarity 19.3%; Pred. No. 0.71;
Matches 106; Conservative 79; Mismatches 198; Indels 166; Gaps 22;

237	QY	TTGVMARFIPSHLNKNSCRVTS	LCYDESGQEIIVSYSSDYIYFDPKDDTARELKT	PSAE	296
9	Db	TTGFFLRSEPT	-----TCRITTFAYSAE	-----IITFHAETTFEFKIMKEGNELEYKSI	59
297	QY	EREELRQPPVKRLRLRGDWSGT	PRARPESRERDGEQSPNVLSMQ	-----RMSDMLSNWF	353
60	Db	ELNSEI	-----WED	-----EEEDNSGIVSQNERLMLKLVNQREI	99
354	QY	EEASEYAQNRGGRSRPRGGT	SQSDISTLPVPVSPDLSEVTAMEVDTPAEQ	FLQPST	413
100	Db	KOLEYKVKONTSLRMRLSKYESTD	FPFSQSPRANSQSDSYSSPYE	-----KGKLFPKISL	156
414	QY	SSTMSAQAHSTSPSTSPHSTP	LLSSPDSQSEQRQSVAEAGSHTHQ	-----SDN	461
157	Db	KS	-----SKDVPITASAH	-----ISSDHEKSSSVLSALANNYKTTDIKARS	205
462	QY	NNEKL	-----SPKPGTGPVLS	-----LHYSTEGT	486
206	Db	TRPKLLNTKSHRSSEEPGASSP	VTSPILKDSQKERTQALRNKAIKTYSV	TESAEPRID	265
487	QY	-----TTSTIKLNTDEWSST	FASSRIGISHCKSEGEESF	-----VPQSSV	538
266	Db	SIRSDNLSPSLNTSSFRPITK	TPPNSSNTSIDPKDNNSNKQDFAETED	ELRQQL	325
529	QY	OPPEGSDTKAPESSSDVTKYQ	GVGAENPVENHINTQDGF	-----	572
326	Db	DIKVRANASSPRKKSISIVK	-PHGISPFKHSNTNLS-SKSGKFHSDFRVVSEN	VLLQAR	383
573	QY	-----TAKPLDNSGERNDLNR	SCCVPEESASSEKAKEPETSDOT	TESATNENNTNPEQ	630
384	Db	SETNSPIIENK-EANFN	-LAPTSNVAYSTPARPESP	-----PPP	423
631	QY	FQTEATGPSAHEETSTRDSALOT	DDSDDDPVLIPCARVYRAGCD	-RRSAVARIQE	685
424	Db	ISSSSTTPRPDKPSP	LPRLSEDNDS-----LSLQKTGSSDTRRSFS	TLKIPDSD	475
686	QY	--FFRRRKE	692		
476	Db	ICFTRRSD	484		

Search completed: March 18, 2003, 15:35:01
Job time : 64 secs

10

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:32:49 ; Search time 23 Seconds
(without alignments)
3594.591 Million cell updates/sec

Title: US-09-781-693A-2

Perfect score: 4527

Sequence: 1 MSRGSGYPHLLWDVRSLSG.....RLEGDRSGSGQENENEDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

First 120 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	13.0	754	2 T50803	hypothetical prote
2	547.5	12.1	809	2 T40574	quanine nucleotide
3	448.5	9.9	481	2 T47502	hypothetical prote
4	429.5	9.5	747	2 S37694	gene PC326 protei
5	405.5	9.0	493	2 T04951	hypothetical prote
6	384.5	8.5	426	2 T05676	hypothetical prote
7	220.5	4.9	489	2 A45988	dentin matrix acid
8	204.5	4.5	1217	2 T00270	hypothetical prote
9	202	4.5	5170	2 T15348	hypothetical prote
10	199	4.4	550	2 T29919	hypothetical prote
11	198.5	4.4	406	2 S38170	SRP40 protein - ye
12	198	4.4	3507	2 T34513	hypothetical prote
13	193	4.3	852	2 T06310	hypothetical prote
14	190.5	4.2	1367	1 S48478	glucan 1,4-alpha-g
15	188	4.2	1702	2 A41859	IgA-specific metal
16	187.5	4.1	2004	2 F95133	immunoglobulin A1
17	186	4.1	1694	2 H64106	IgA-specific metal
18	183	4.0	727	2 T29612	hypothetical prote
19	183	4.0	2271	2 F90073	hypothetical prote
20	180.5	4.0	868	2 F89500	hypothetical prote
21	178	3.9	1027	2 T46481	protein T27A10.6
22	178	3.9	1051	2 T48933	hypothetical prote
23	177.5	3.9	1063	2 S51101	WD repeat domain p
24	177.5	3.9	1658	2 S51101	hypothetical prote
25	176.5	3.9	734	2 B42680	hypothetical prote
26	176.5	3.9	802	2 T21315	nucleolus-cytoplas
27	176.5	3.9	5327	2 T13564	hypothetical prote
28	176	3.9	1459	2 T32271	microtubule-associ
29	175.5	3.9	1849	2 C41859	IgA-specific metal

30	175	3.9	534	2 T39903	serine-rich protei
31	175	3.9	1063	2 T38732	probable helicase
32	175	3.9	1829	2 T24583	hypothetical prote
33	174.5	3.9	787	2 A48819	nuclear autoantige
34	174.5	3.9	1587	2 G86467	hypothetical prote
35	174	3.8	1200	2 A46194	neurofilament prot
36	174	3.8	2232	2 T34434	hypothetical prote
37	173	3.8	895	2 T45738	hypothetical prote
38	173	3.8	1641	2 T38614	helicase II - huma
39	173	3.8	1963	2 B98002	IgA-specific metal
40	172.5	3.8	1233	2 T30989	serine/threonine p
41	172	3.8	664	2 S60062	hevin precursor -
42	172	3.8	1007	2 T01437	hypothetical prote
43	171.5	3.8	1337	2 T30291	dextranase - Strep
44	170.5	3.8	1189	2 S56852	hypothetical prote
45	169.5	3.7	619	2 JX0092	apopolysialoglycop
46	169.5	3.7	1092	1 JN0635	neural cell adhesi
47	169.5	3.7	1589	2 C44766	defective chorion-
48	168	3.7	910	1 S73361	dnaj homolog prote
49	167.5	3.7	2342	2 T13412	hypothetical prote
50	166.5	3.7	1359	2 T34036	hypothetical prote
51	166.5	3.7	2722	2 T20532	hypothetical prote
52	166	3.7	643	2 T19199	hypothetical prote
53	165.5	3.7	891	2 G84693	probable proline-r
54	165.5	3.7	1210	2 T39410	AF-4 protein, spli
55	165.5	3.7	1435	2 A37793	erythrocyte-bindin
56	165.5	3.7	3924	2 S37431	ankyrin 2, neurona
57	164.5	3.6	660	2 JMW0067	chitinase (EC 3.2.
58	164.5	3.6	1419	2 T30531	agglutinin-like ad
59	164	3.6	1188	2 T05324	hypothetical prote
60	163.5	3.6	971	2 T19431	hypothetical prote
61	163.5	3.6	1747	2 AC1842	WD-40 repeat prote
62	163	3.6	1466	2 A36426	SPA2 protein - yea
63	163	3.6	6842	2 T29757	protein UNC-89 - C
64	162.5	3.6	699	2 T38073	nucleolar phosphop
65	162.5	3.6	1213	2 A58198	serine/proline-ric
66	162.5	3.6	1630	2 A53577	ascites sialoglyco
67	162	3.6	727	2 T24284	hypothetical prote
68	162	3.6	1032	2 T34433	hypothetical prote
69	161.5	3.6	671	2 S53407	CHS5 protein - yea
70	161	3.6	1390	2 T42004	trfa protein - sli
71	161	3.6	3938	2 T42761	Bassoon protein -
72	160.5	3.5	1072	2 A86827	hypothetical prote
73	160.5	3.5	1175	2 T25634	hypothetical prote
74	160.5	3.5	1356	2 T18521	beta transducin-li
75	160.5	3.5	4377	2 A55575	ankyrin 3, long sp
76	159.5	3.5	1386	2 T49316	profilaggrin relat
77	159.5	3.5	1560	2 T30282	calcium-binding pr
78	159.5	3.5	2481	2 D90011	FmtB protein (limp
79	159	3.5	961	2 E86245	hypothetical prote
80	159	3.5	1275	2 T38397	probable GTPase ac
81	159	3.5	1310	2 T40135	probable involveme
82	159	3.5	1711	2 AD1842	WD-40 repeat prote
83	158.5	3.5	1104	2 S59310	probable membrane
84	158.5	3.5	2738	2 E88320	protein F07A11.6
85	158	3.5	1280	2 T00365	hypothetical prote
86	158	3.5	680	2 A43359	microtubule-associ
87	157.5	3.5	1912	2 A43800	nuclear autoantige
88	157.5	3.5	1151	2 T29088	vitellogenin I pre
89	157	3.5	1286	2 T33777	hypothetical prote
90	157	3.5	1615	2 A88396	protein M01E10.2
91	157	3.5	532	2 JC6510	ras-responsive ele
92	156.5	3.5	713	2 T06029	hypothetical prote
93	156.5	3.5	3942	2 JN0133	WD-40 repeat regul
94	156.5	3.5	1015	2 T42730	Bassoon protein -
95	156	3.4	1211	2 S68141	nuclear protein HI
96	156	3.4	1791	2 T42230	AF4 protein - mous
97	156	3.4	2957	2 T02345	hypothetical prote
98	156	3.4	505	2 T33152	hypothetical prote
99	155.5	3.4	786	2 B64560	poly E-rich protei
100	155.5	3.4	1020	2 T16509	hypothetical prote
101	155.5	3.4	1217	1 QFHDH	neurofilament trip
102	155.5	3.4		2 T42625	AF-4 protein - mou

103 155 3.4 1062 2 S61196
104 155 3.4 1131 2 T41144
105 155 3.4 1229 2 T25697
106 154.5 3.4 1299 2 A25680
107 154.5 3.4 1095 2 T43275
108 154.5 3.4 1684 2 JW0057
109 154.5 3.4 2218 2 B84683
110 154 3.4 1093 2 A47212
111 154 3.4 2938 2 T30249
112 153.5 3.4 519 2 S45345
113 153.5 3.4 831 2 T48442
114 153.5 3.4 913 2 T52485
115 153 3.4 699 2 T01029
116 153 3.4 725 2 A41258
117 153 3.4 888 2 T46726
118 153 3.4 1192 2 T18611
119 153 3.4 2526 2 T20531
120 152.5 3.4 606 2 A43427

SUM1 protein - yea
hypothetical serin
hypothetical prote
nuclear histone-bi
neurabin - rat
gravin - human
hypothetical prote
transcription fact
cell proliferation
Tup1-like enhancer
hypothetical prote
neurofilament prot
hypothetical prote
a-agglutinin core
secreted acid phos
probable serine/th
hypothetical prote
neurofilament trip

RESULT 1
T50803
C:Species: Arabidopsis thaliana
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T50803
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mey
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25240
A:Accession: T50803
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-754 <BEV>
A:Cross-references: EMBL:AL365234
A:Experimental source: cultivar Columbia; BAC clone T30N20

ALIGNMENTS

Query Match 13.0%; Score 590.5; DB 2; Length 754;
Best Local Similarity 24.6%; Pred. No. 2.9e-24;
Matches 207; Conservative 114; Mismatches 279; Indels 245; Gaps 28;

QY 22 BDPS-RLRSRYLGRREFTORLKEATLVNHDGCVVTCWNTGCEYILSGSDTKLVISNP 80
DB 20 QDPSHEVDQMQLHSLVRLSQEQLGHCQVNALAWNSGLLSGSDDLRLINWY 79
QY 81 YSRKVLTVIRSGHRANIFSAKFLPCTNDKQIVSCGDGVI--FYTN-----VEQDAETNR 133
DB 80 SSRKLLHSIDTGTANIFCTKVPETSDLVVSGAGDAEVRLENTSLSGRAEDDNLAIIP 139
QY 134 QCQFTCHYGTVEIMTPNDPFTYILSCGREDGVRVWFDPRIKTSK-----TKEDCKDDILL 188
DB 140 SALYQCHTRRVKKLAVEPCGNPNVWWSASEDGLRQHDRESTSCPPAGTAHQECR-SVLL 198
QY 189 NCRRAATVAICPP-----IPYVLAVGCSDDSVRIYDRMLGTRATG-----230
DB 199 DLRSAGKALADPPKOTLSLKSDISATRPHLLLVGGSDAFARLYDRMLPPLASSRKM 258
QY 231 -----NYAGRTGTGMVAFIPSHLNK-----SCRVTSLCYSDGQELVSYSDYILFD 281
DB 259 PPPPCVNY-----FCPMHLSRGRTNLHLTHVTSPNGEELVLSYSGHERVYLMN 307
QY 282 PKDDTARELTPSAERERELRPPVKRLRLRGDWSDTGPRAPRERERDQSPNVSL 341
DB 308 VNGTGMQYTPGDVDNLFSE-----SNLHDSVPPQVS-342
QY 342 NQMSDMLSRWEASEVAQNGRGRSRPRGGTQSOSDITSLPTVPSPDL-EVSETAME 400

343 -----TTPQNGFHS--SNAATVKKCTELVEIAKWSLE 373
401 VDT-----PAEQFLQSTSTMSAQAHSTSSP-----TESPHSTPLSPDSE 443
374 EGTDFVYAEAAANEVLDAHSDIESALRECLCTRAALLKKRWKNDH-----MAVRDCH 429
444 QRSVEASGHHTHQSDDNNNEKLSKPKGTGEPVLSLHYSTEGTTTSTIKLNFIDEWSSIA 503
430 NARRIDASSFKAHYIMSEALQOL-----GKCEALDFATAAHMNPSPADIYAKVESIK 483
504 SSSRGIGSCKEGQESFVQSSVQPPGDSSTK-----APESSSEDVTKYQ 551
484 RDLQAAG-----AEKNEET-----GAGTRVLSLSLILYRSEANSSDSDHMSRSE 528
552 -EGVSAENPVENHINTQSKDTAKPLDLSNGSRNDLNRSCGVPPEESASSKAKEPET 610
529 REDSDYDEELELDIQISLSD--EGRDTDSNS--MRGSLNL-----RIHVGDDKPMK-NT 579
611 SDQTSFESATNENNTNPEQFOTEPATGSAHETSTRDSALQDTSDDDDPVLIPGARYR 670
580 VDNASSGTASSQN-----DRTSYQPEGAID-----605
606 -----MKRRYVGHCVGTDIKQASFLG 627
731 --ANFVMSGDCGCHIFTWDRHTAEHMLLEADNVVNCLOPHPPDPILASSGIDYDIKI 788
628 QRGEYIASGSDGGRWFIEWKOTGRIMKLVGDESVLNCIOCHPPFDSVVATSGIDNTIKI 687
789 SP 790
688 SP 689

RESULT 2
T40574
guanine nucleotide binding protein beta subunit-like - fission yeast (Schizosaccharom
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40574
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21938
A:Accession: T40574
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-809 <SEE>
A:Cross-references: EMBL:AL035226; PIDN:CAA22832.1; GSPDB:GN00067
A:Experimental source: strain 972h-; cosmid c609
C:Genetics:
A:Gene: SPAC609.03
A:Map position: 2
A:Introns: 21/1; 52/2; 280/3; 780/2

Query Match 12.1%; Score 547.5; DB 2; Length 809;
Best Local Similarity 24.8%; Pred. No. 6.4e-22;
Matches 236; Conservative 124; Mismatches 308; Indels 285; Gaps 41;

QY 27 LRSRYLGRREFTOR-----LKLKLEATLVNHDGCVVTCWNTGCEYILSGSD 71
DB 6 LSLRLQDYRDWFQKISRDYGNSTWLTGIDLQKELTGTCVNTLWDSADGEFLSGSD 65
QY 72 DTKLV---ISNPYSRKVLTVIRSGHRANIFSAKFLPCTNDKQIVSCGDGVIYFTVNEQD 128
DB 66 DTRLIWVDVFNKKPHRL--ISGHVQNIQFSAKFPVYNNRQILSLASGDKLKLFDSS 123
QY 129 -----AFTNQCOFTCHYGTVEIMTPNDPFTYILSCGREDGVRVWFDPRIKTSCKE 180
DB 124 KEGMDHGMETQTRC-WSCALDSVKNIIVPCDNG-HTFLVCSDEGTARQYDIREFHVCNQD 181
QY 181 -DCKDDILLIN--CRRRAATVAICPPYIPYVLAVGCSDDSVRIYDRMLGTRATG-----TGNY 232

Db 182 LDC-PSILVNPYRINLYTITMSPNPPYFAIGGTHPYAFLYDRRMVKKSPRDDWTMT 240
 QY 233 AGRGTGMVAFIPSHLNNKSC-----RVTSLEY--SEDOEILVSYSSDYIYLFDPK 283
 Db 241 SPEKDCRCVRKFSF-----DGCSQSGILDRTYTCQFSANPNELLVSNDSYIYLFVHV 296
 QY 284 DDTARELTPSAEERREELRQPPVKRLRGDWSDTGPRAPESRERDGEQSPNVSLMQ 343
 Db 297 EDKS---YTPFNKEDSNKKPSKPSL-----LQTOPLKRN----- 330
 QY 344 RMSDLSRWFEEASEVAQSENRGRGRSRPRGTSQSDISTLTPTVSPSPLEVSETAMEVDT 403
 Db 331 -----YSPWY-----KNNFG-----AST 343
 QY 404 PAEQFLQSTSTMSAQHSTSTSPESHSTPLLSPPDSEQRQSVASGHHTHQSDNNN 463
 Db 344 PA-----SRVSNPYYTAAQPK-----HTFQMYENI 370
 QY 464 EKL--SPKPGTGEPLVSLHYSTEGTTSTIK--LNEFTDWSIASSS----- 506
 Db 371 EKFTTENGGLYESIVGRLSHFSRSIQYKDAIFYLENYNIYIPDSNGLNHSIRVSALRY 430
 QY 507 -RGIGSHCKSEGQESFVPOSSVQPPGDSKAPESSESDVTYKQEGVS-----A 556
 Db 431 WRACVSLALMDDTVSLSPNTIIQAGWG-----WLYDFMNVWTRYLLGSDHWALQMSPP 485
 QY 557 ENPVENIINTQSKDTAKPLDLSNGE-----RNDLN-----LDRSCGYPE 597
 Db 486 TNVARQNFVLCDDPE-PSRVLFNPSMIRAFARIDTNDLSRRRFRFVHKLVRGCLLI 544
 QY 598 ESASSEKAKEP---ETSDQTSATNENN---TNPEQFQTEATGPS-----AH 641
 Db 545 SSDIYWEQFQPMDSSTSDVTSSQRLDDENGFLTLLEPPYNYENEVSSEGENIVSMYTG 604
 QY 642 EETSTDSALQD-----TDSDDDPVLPAGARYRAGPDRRSARVARIQEFF 687
 Db 605 SLDNDDDDYODEESYASDDDESDSDSDEGPTLLS----- 642
 QY 688 RRKEREKMEELDNLIRRPVLMVYKGRHNSRTMKLEAFWGAN--FVMSGDCGHIFI 745
 Db 643 LRMKRKAQVPPVNVTH---VKSYY-GHCNVES-IKNVFTYQNDYEVMSGSDGGRFFI 697
 QY 746 WDHHTAHLMLLEADNHVNCVLPHPDPILASSGIDYDIKIWS-----PLESRIFNRK 800
 Db 698 WKLNASILAIHGDSEAVNYTEGHPRCTPLAVSGIDSTVKIFNTENTPPSGSRNHTSN 757
 QY 801 LADEVITRNLMELETRNTITVPASFLMRMLASLHNRADLRGDRSESGQE 853
 Db 758 -SYKIATNEMNRQGRSDSYI-TSRMLSHLAYRAHL-----DDGFGHE 799

RESULT 3
 T47502
 Hypothetical protein F9K21.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T47502
 R:Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Lemcke, K.
 submitted to the Protein Sequence Database, February 2000
 A:Reference number: 224467
 A:Accession: T47502
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-481 <J0R>
 A:Cross-references: EMBL:AL138657
 A:Experimental source: cultivar Columbia; BAC clone F9K21
 C:Genetics:
 A:Map position: 3
 A:Introns: 39/3; 123/3; 171/3; 406/3
 A:Note: F9K21.200

Query Match 9.9%; Score 448.5; DB 2; Length 481;
 Best Local Similarity 19.2%; Pred. No. 6.7e-17;

Matches 167; Conservative 72; Mismatches 194; Indels 437; Gaps 16;
 QY 13 DYKRSLGLEDPSRLRSRYLGRREEFORLKLKLEATLVNHGCVNTICWDTGETYILSGSD 72
 Db 17 EIFNEIGFSPHITTSRISASAGRVKKLDLYLKGUNGHEGCVNAVEFSTGDLVSGSD 76
 QY 73 TKLVLSNPYSRKVLTIIRSGHRANIFSAKFLPCTNDKQIVSCGDCGVITVTVNED--AE 130
 Db 77 RQIMLNMWLSGRKLSYPSGHCENFQTKFIPFDDRTIITSGADGQVPLGILENGKVE 136
 QY 131 TNROQFTCHVGTVEIMTVNDPYFLSCGEDGVIRVDFTRIKTSCKEDCKDDILINC 190
 Db 137 TKRLGR---HGRVYKLVLPDGNVNFYSCGDBGVQVHFDIR-SNSATMVLVSSPFTQGC 192
 QY 191 RR-----AATVAICPPPIPYLAVGSDSSVRYIDRRMLGTRATGNVAGRTTGMVAR 243
 Db 193 RRHSSSRIRLNSAIDPRNSYILAVGSGDEVARVYDTRVQLAPVCRHV--LPDAPVNT 250
 QY 244 FIPSHL-NKSCRVTSCLYSEDGQELVSYSDYIYLFDPKDDTARELKTTPSAERREEL 302
 Db 251 FCPHLRETNSVHITGLAYSAG-ELLVYNDELIYLF----- 287
 QY 303 RQPPVKRLRLRGDMSDTGPRAPESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQS 362
 Db 288 -----EK 289
 QY 363 NRGGRSRPRGGTSQSDISTLTPTVSPSPDLEVSETAMEVDTPAEQFLQPSSTSTMSAQAH 422
 Db 290 NMGYG----- 294
 QY 423 STSPSTESPHSTPLLSPPDSQRQSVASGHHTHQSDNNNEKLSPRPGTGPVLSLHYS 482
 Db 295 --SSPV----- 298
 QY 483 TEGTTTSTIKLNFTEWSSIASSRGIGSHCKSEGOESFVQSSVQPPGDSKAPTEE 542
 Db 299 ----- 298
 QY 543 SSEDVTYQBGVAENPVENHINTQSDKFTAKPLDSNGERNDNLDRSCGVPEESASS 602
 Db 299 -----SVSP 302
 QY 603 EKAKEPETSQDSTESATNENNTNPQFQTEATGPSAHEETSTDSALQDTSDDDPV 662
 Db 303 EKLQEME-----EPQ----- 312
 QY 663 LIPGARYRAGPDRRSARVARIQEFFRRRKRKEMEELDTLNRPLVKMYKGRHNSRTM 722
 Db 313 -----VYIGHRNAQT- 322
 QY 723 IKEANFWGAN--FVMSGDCGHIFWDRHTAHLMLLEADNHVNCVLPHPDPILASSG 780
 Db 323 VKGVNFFGPNDEYVTVSGDCGHIFWKKGGKLVRAWVGDRRVNVNQLSEHPHPLLASCG 382
 QY 781 IDYDIKIWSPLEESRIFNKKLADEVITRNLMELETRNTITVPASFM--LRMLASLHNR 838
 Db 383 IEKSVKLTPTMSNDVLSLEPKIDKVMELNRVGRDQSRQSVTLTPDVIMVLRQRRTSAP 442
 QY 839 ADRLGDRSESGQENE-----NEDE 859
 Db 443 TERRVSTDTIGSDEGNDARFTIASLVANDDE 472

RESULT 4
 S37694
 gene PC326 protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000
 C:Accession: S37694; S27875
 R:Bragsagel, P.L.; Timblin, C.R.; Eckhardt, L.; Laskov, R.; Kuehl, W.M.
 Oncogene 7, 2059-2064, 1992
 A:Title: Sequence and expression of a murine cDNA encoding PC326, a novel gene expres
 A:Reference number: S37694; MUID:93026383; PMID:11406147

A:Accession: S37694

A:Molecule type: mRNA

A:Residues: 1-747 <ECK>

A:Cross-references: EMBL:M95564; NID:g200240; PIDN:AAA39895.1; PID:g200241

C:Genetics:

A:Gene: PC326

C:Superfamily: unassigned WD repeat proteins; WD repeat homology

F:349-382/Domain: WD repeat homology <WD1>

Query Match 9.0%; Score 429.5; DB 2; Length 747;
Best Local Similarity 18.7%; Pred. No. 1.2e-15;
Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

QY 8 PHLLMDV-----RRKSLGLDEPSRLSRYLGRREFIQRKLEATLNVDGCVNTICNDTG 63
DB 308 PRPRNVLNLRDLQGL--SSGRFVYACCARLFVQRFSLFVFEHSGCGVNTVHFNQHG 365
QY 64 EYILSGDDTKLVISNPYSRKVLTTRSGHRANIFSAKFLPCTNDKQIVSCSDGQVIFYT 123
DB 366 TLLASGSDDLKVIWDLKRSVNTDSGHKNILQAKFLPNCNDAILAMCGRDGQVRVA 425
QY 124 NVQDAETNRCQOFTCHYGTTEIMVPNDPYTLSCGEGDTGVWFDTRIKTSCTREDCK 183
DB 426 QLSAVAGTHMTKRLVKHGGASHRLGLEPDSFPRFLTSGEDAVVFNIDLRQAHFASK 481
QY 184 DDILI-----NCRRAATVAICPPPIPYLVAGCSDSSVRIYDRMLGTRATGNTAGRTTG 239
DB 482 --LLVIKDGDKKVLGYTVFVNPNVYQFVGGQDFRIYDQKIDENV-----NNG 531
QY 240 MVARFIPSHL--NNKSCRVTSLCYSEDQGEILVSYSDIYLFDPKDDTARELKTPTSAAE 297
DB 532 VLKFCPHLLSSDPYAHITSLMYSYDGTETILASYNEDIYF----- 574
QY 298 RRELRPPVKRLRLGDSWDTGPRAPESERERDGEQSPNVSLMQMSDMLSRWFEEAS 357
DB 575 ----- 574
QY 358 EVAQSNRGRSRPRGTSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLOPSTSTSM 417
DB 575 ----- 574
QY 418 SAQAHSTSPSPESHSTPLLSPPDSEQSQVSEASGHHTHQSDNNNEKLSKPGKGPVPL 477
DB 575 ----- 574
QY 478 SLHYSTEGTTTSTIKLNFTEWSSIASSSRGIGSHCKSEGESEFVQSSVOPPEGDSET 537
DB 575 ----- 574
QY 538 KAPEESEDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGSGERNDLNLDRSCGVPE 597
DB 575 ----- 574
QY 598 ESASSEKAKEPETSQSTESATNENTNPEQFQTEATGPSAHEETSTRDSALQDSTD 657
DB 575 -----NSSDS 579
QY 658 DDDPVLIPGARYRAGDRRSARVARIQEFRRRKRKEMEELDTLNRPLVKNYKGRH 717
DB 580 D-----CAQY-----AKR-----YKGRH 592
QY 718 NSRTMKENFWG--ANFVMSGDCGHIFIDWRHTAHLMLLEAD--NHVVCILQHPFDP 774
DB 593 NNST-VKGVYFYGPRSEFVMSGDCGHIFVWKSQCIQVQFLEADGGTINCIDSHPLP 651
QY 775 ILASSGIDYDIKIPLESRIFNKLA--DEVITNEMLEB--TRNTITVPAS---FM 827
DB 652 VLASSGLDHEVKIWSPIAEP---SKKLAGLKNVINKLKRDNFTLRHTSLFNNSMLCFL 708
QY 828 LRMASLNH-----IADRLGDRSGSGQENENEDE 859
DB 709 MSHVTSQNTGRSWRGIRINAGGGDFSDSSSSSEETNOE 746

RESULT 5

T04961

hypothetical protein T12J5.10 - Arabidopsis thaliana

N:Alternate names: hypothetical protein M4E13.190

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999

C:Accession: T04961; T05784

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Hoheisel, J.; Mewes, H.W.; Mayer, K.

submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15392

A:Accession: T04961

A:Molecule type: DNA

A:Residues: 1-493 <BEV>

A:Cross-references: EMBL:AL035522

A:Experimental source: cultivar Columbia; BAC clone T12J5

R:Bevan, M.; Purnelle, B.; Boutry, M.; Goffeau, A.; Hoheisel, J.; Mewes, H.W.; Mayer, K.

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15451

A:Accession: T05784

A:Molecule type: DNA

A:Residues: 398-493 <BEW>

A:Cross-references: EMBL:AL022023

A:Experimental source: cultivar Columbia; BAC clone M4E13

C:Genetics:

A:Map position: 4

A:Introns: 43/3; 127/3; 371/3

A:Note: T12J5.10; M4E13.190

Query Match 9.0%; Score 405.5; DB 2; Length 493;

Best Local Similarity 18.9%; Pred. No. 1.4e-14;

Matches 162; Conservative 74; Mismatches 151; Indels 471; Gaps 20;

QY 9 HLLMDVKKRSGLDEPSRLSRYLGRREFIQRKLEATLNVDGCVNTICNDTG 68
DB 17 HPVDFWRFRGSGSSRNFSASENLVLRLEIKLKHKGCVNTVFNAGDVLIS 76
QY 69 GSDTKLVISNPYSRKVLTTRSGHRANIFSAKFLPCTNDKQIVSCSDGQVIFYTNNVED 128
DB 77 GSDRVRVLDWQIOLGNVLSFHSGHANNVFOAKPMFSDDRITVTCADGMF---DLRTE 133
QY 129 AETNRCQOFTCHYGTTEIMVPNDPYTLSCGEGDTGVWFDTRIKTSCTREDCKDDILI 188
DB 134 APTE---LFTC-----RSVDPRRN-----MDATQL 156
QY 189 NCRRAATVAICPPPIPYLVAGCSDSSVRIYDRMLGTRATGNTAGRTGMVA---RFI 245
DB 157 N-----ALADPRNSNLFVGGMEYARLYDIR-----RFQEGGLNGFTRAADHFC 202
QY 246 PSHL--NNKSCRVTSLCYSEDQGEILVSYSDIYLFDPKDDTARELKTPTSAAERRELRQ 304
DB 203 PPHLLIGNEDVGTGLAFSEQ--SELLVSYNDEFIYLFTP----- 239
QY 305 PVKRLRLGDSWDTGPRAPESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNR 364
DB 240 ----- 239
QY 365 GRGRSRPRGTSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLOPSTSTMSAQAHST 424
DB 240 GMGLG----- 244
QY 425 SSTESPHSTPLLSPPDSEQSQVSEASGHHTHQSDNNNEKLSKPGKGPVLSLHYSTE 484
DB 245 -----SNIPSPSPIS----- 254
QY 485 GTTSTIKLNFTEWSSIASSSRGIGSHCKSEGESEFVQSSVOPPEGDSETKAPEES 544
DB 255 -----KSPVSKS 261
QY 545 EDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGSGERNDLNLDRSCGVPEESASEK 604
DB 262 -----ESSSPK 268

QY	32	LGRRFIQRLKLEATLNVDGCVNTICWDDTGEYILSGSDDTKLVTSNPSYRKVLITWRS	91
Db	3	: : : : : : : : : : : :	1
QY	3	LQKODLLRRLGDKKDKHKGCVNTVSFNADGILLSGSDDRQVILWDQOTASVKLSFDS	62
Db	92	GHRANIFSAKFLPCTNDKQIVSGSGDGVIFYTNVEDQATFNQCQFCHVGYTYEITWVP	151
Db	63	GHPNNIFQAKFMFSDRITVISAADQVRYSKILESGQVETSL-IGKHQGPVHKLAVEP	121
QY	152	NDPYTILSGCGDGTVR-----WDFRIKTSCTKECDKDIINCRR-----AT	195
Db	122	GSPFSYTCGEDGAVKHVSCIFHFVPLSTRVATN-----LFTCKEAFNLVYLH	171
QY	196	SVAICPEPIYLLAVGCSDSRVIRYDERMLGTRATGNVACGGTGMVARFIPSHLNKSCR	255
Db	172	ATAVDPNGLLAVAGMDYFARVDIR--SYRSEDHVG-----	208
QY	256	VTSLCYSEDGQELVSYSDYLYLDPKDDTARELKTPSAERREBELQPPVKRLRLRGD	315
Db	209	ITGLAFS-DOSELLASYSEFIYLTTP-----	234
QY	316	WSDTGPRAPSERERDGEQSPNVSLMORMSMLSRWFEAEVAQSNRGSRSPRGCT	375
Db	235	-----DM-----GLG-----	239
QY	376	QSQSDITLTPVSSPDLEWSETAMEVDTPAEQFLQPSSTSMAQAHSSTSPTESPHST	435
Db	240	-----PTPYPSSTK	248
QY	436	LLSSPDSBQRQSVASGHHTHHQSDNNNEKLSPPGCTGEPVLSLHYSTEGTTSTIKLNF	495

Query Match	4.9%;	Score 220.5;	DB 2;	Length 489;
Best Local Similarity	25.9%;	Pred. No. 0.00012;		
Matches	107			

Query Match	4.9%;	Score	220.5;	DB	2;	Length	489;
Best Local Similarity	25.9%;	Pred. No.	0.00012;				
Matches	103;	Conservative	55;	Mismatches	165;	Indels	75;
Gaps							
QY	281	DPKDDT-ARELTPSAERREELRPVVKRLRLRGDWSDTGPRAPRSESRDGPSPNV	339				
DB							
DB	89	DSGDDTFGEDNNGPPEER----	QWGGPKRLSDSDSADTTQSSDSTQENSAQDTFSD	144			
QY	340	SLMQMSDMLSRWFEEASVAVO--	SNRCGRGRPRGGTQSQDSTLTPVPSPPLEVSE	396			
DB							
DB	145	SKDHDSDEADR--PAGDSTQDSEEEYRVGGSEGSFHGDG----	EFDD	191			
QY	397	TAMEVTPAEQFLQPTSTSTMSAQAHSTSSPT----	ESPHSTPLLLSPDPSRQSRQSVFAS	451			
DB							
DB	192	EGMQSDDP-----	GSTRSDRGHTRMSSADISSEFSKGDHETPTSQDSDSDQVFEFS	242			
QY	452	GHHTHQSDNNNEKLSPKPGCTCEPVLSHYSEGHTTTSTIKLNFTEWSSIASRSGIGS	511				
DB							
DB	243	SRKSPFRS-----	RVSEEDDRGELADNSRSTQSVSTEDFRSK-----	E	81		
QY	512	HCKSEGESEFVPQSSVOPPEG-----	DSEWKAPESSDVTYKQEGVSAENPVENHINI	566			
DB							
DB	282	ESRSETQEDTATQOQDQSPQEQDPSSESEAEQPPQSSSESEQEGVASESRGNDPNT	341				
QY	567	TQSDKFTAKPLDNSGERNDLNLDRSCGVPPESASSEKAKETPSD-----	QTSESATN	621			
DB							
DB	342	SO-----	TGQPRDSSESSEDRLNTFS-----	SESQSTEQGDSSESNESLSUSEEQESAQD	393		
QY	622	ENNTNPEP-QQQTATGSAHEHTSTRDSALQDITDSD	658				

Db 394 EDSOQGLQSQASRESRSQESQSEDSRSRENDS 431

RESULT 8

T00270

hypothetical protein KIAA0596 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00270

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete

A:Reference number: Z14086; PMID:98290545; PMID:9628581

A:Accession: T00270

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1217 <NAG>

A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA25522.1; PID:g3043716

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0596

Query Match 4.5%; Score 204.5; DB 2; Length 1217;

Best Local Similarity 20.6%; Pred. No. 0.0027;

Matches 187; Conservative 121; Mismatches 351; Indels 247; Gaps 43;

QY 39 QRLKLEATLVHDCVNTICWN--DTGEVILSGSDTKLV-----ISNPYSRKVLTTIRSG 92

Db 218 EMLKEA-----HDEILGLEYKPTGTLKLKASRDLIHVLDAGRYIS---LQOTLDE 270

QY 93 HRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDQAEFTNQCQFT-CHY-----GTTVEI 147

Db 271 HSSITATVKAASDQGVRMISGADKSIYFRTAOKSGD---GVQFTRTHHVVRKTKLYDM 327

QY 148 MTPVNDPTFLSCGDEGVTRVFDRIKTSCTKEDCK-----DDILNCRRRAATVAICPP 202

Db 328 DVEPSWKYTAIGC-QDRNIRIFN--ISSGOKKLFKSGDEGTLIKVQ-----TDP 376

QY 203 IPYLVAGCSQSSSVRIYDRMLGTRATGNAGRGVYTGVMARFIPSHLNKSCRVTSLSYS 262

Db 377 SCIIYATSCDKNLSIFD-----FSSGECVATMFGH-----SEIVTCMKFS 417

QY 263 EDQGEILVSYSD-YIYLPDKDDTARELKTPTSAEREE---LRQPPVKRLRLRG--- 314

Db 418 NDKCH-LISVSGDCIFVWLSEMTISMQRQLAELRQGRKGQGGQSPSPORASGPNRH 476

QY 315 ---DWSDTGPRAPRESERDQ---EQSPNVSLMQR-----MSDMLSRWFE 354

Db 477 QAPSMSPGALSSDSDKEGEDETEELPALPVLAKTAKALASVPSPALPRSLHW-- 534

QY 355 EASEVAQS-----NNGRGRSRPRG-----GTSQSD 379

Db 535 EMSRAQSGVGLDPAPANPGRRRGRVQVQVGLSVRGMLDLRLQETLAPSLQDPSQDS 594

QY 380 ISTLPVSPDLEVSETAMEVDTPAEQFLQSTSTMSAQAHSTSPSTESPHSTPLLS 439

Db 595 LATIPSGPKRHQGEALETSL-----TSQNEKPPRQAQPCSYPIIRLLSQ 641

QY 440 PDSEQRQSVAS---GHHTHHQSDN-----NNEKLSPK 469

Db 642 EGVFAQDLPAFIEDGIYVPEPDNDTMDTSEFQVQAPARGTLGRVYPGSRSEKHP- 700

QY 470 PGVGEVLSLHYTEGTTSTIKLNTDEWSSIASRSGISHCK--SEGOESEFVQSS 527

Db 701 ----DSACSVDYSS--SCUSSPHEPTDESETEPLSDVGLSSDLEPAPAEDEEEEGG 754

QY 528 VQP---PEGDSETKAPE--SEEDVTKYQGVSAENPY---ENHINITQSKF-----TAKP 576

Db 755 MGYGLQEGSPQIPDQGLKHFETLASGAAPGVQVPERSESRISIRFLIQVQTRP 814

QY 577 LDNSGGRNDLNL--DRSCGVPEESASSEK---AKEPETSQDTSTESATNNTNPEQ-- 630

Db 815 LREPSPSSSSLALMSRPAQVQASGEQPRNGANPPGAPPEVEPSSG-----NPSFQA 868

QY 631 -----FQTEATGFSAEHETSTRDSALQDSTDSDDDVPLIPGARYA-----GPDGR 676

Db 869 ASVLLPCRNLNPDSSWAPKRVATASPSG:KQAQSVHS---LVPQERHEASLOAPSPGAL 925

QY 677 RSARVARIQEFRRRKERKEMEELDTLNRPLVMYVKGH-----RNSRTMLKEANFWG 730

Db 926 LS-----RETEAQDGLSLPPADGPPSRPHSYQNPTTSSMAKISRISVSG 970

QY 731 ANFVMSGSDCGHIFINDRHTAEHLMLLEADNH--VNCLOQPHPPDPILLASSGIDYDIKIWS 789

Db 971 ENGLVAEPAQHAPI--RVSPLSKIALPSRAHLVLDIPKPLDRPTLAA-----FS 1019

QY 790 PLEESR 795

Db 1020 PVTGKR 1025

RESULT 9

T15348

hypothetical protein B0350.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15348

R:Gattung, S.

A:Submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid B0350.

A:Reference number: Z18332

A:Accession: T15348

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5170 <GAT>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1; CESP:B0

C:Genetics:

A:Gene: CESP:B0350.1

A:Introns: 48/1; 5039/3; 5116/3

Query Match 4.5%; Score 202; DB 2; Length 5170;

Best Local Similarity 20.7%; Pred. No. 0.024;

Matches 125; Conservative 84; Mismatches 248; Indels 146; Gaps 24;

QY 191 RRAATSVACIPPIPYLVAGCSDSSVRIYDRMLGTRATGNAGRGVYTGVMARFIPSHLN 250

Db 520 RPISSEPAVSEREPHIAETTTTTRDYDQEPKADQSSARES-----VRDLVSEEH 572

QY 251 NKSCRVTSICYSEDGQEIIVSVS-SDYIYLFDPKDDTA--RELKTPSAER-----REE 301

Db 573 HPACDSERLSEPAQSPPEPVETHAESQFSLVETTTTAVTREFYDEDEQEOASSRATKER 632

QY 302 LRQPPVKRLRLRGDWSDTGPRARP--ESERERDGEQSPVSLMORMSMLSRWFE--AS 357

Db 633 IEQSPVASER---SIVSTEHRSQPTQESQSLPTSEKNVHTVTETTTVTRCFEPPIAS 689

QY 358 EVAQSNR--GRGSRPRGGTSQSDISTLTPVPSPP--DLEVSETAMEVDTPAEQFLQPS 414

Db 690 ELBHAREFEQGNDRFRSVSDVADQEVDPHSPASHAESEQPVPEPHHIVETTTI 749

QY 415 STMSAQAHSTSSP-TESPHSTPLLSGPDSPQRQ-----SVEASGHTTHQSDNNN 463

Db 750 TTVTRFQDEEYPRPESP--AEIFFPISPSQSEQSEPHIVKETTITTTVTRRELDEPEKGN 807

QY 464 EKLSRPGT-----GEPVLS---LHSTEGTTSTIKLNTDEWSSIASRSGIGS 511

Db 808 VTFSPAPSSHAESERQVPSPVVSHQYPPHVTITTTNTVNTSNLYDDEDNVPSSEDPATQ 867

QY 512 HCKSEGESESVFQGS---SVQPPPEGDS-----SVQPPPEGDS----- 535

Db 868 HFQ---QSETSVHRSHDPSVEESDGEGLGSKVLGFAKKAGMAGVGVAAAPVALAAGAKA 924

QY 536 -----ETKAPESSEEDYTKYQF--GVSAENPVENHINITQSKDKTAK 575

Db 925 AYDAFEKDEDEDETSHSPSPVPEYQSE---QYQDDSAQSSHTDFEHM----- 970

QY 576 PLDSNGERDNLDRSCGVPESASSEKAKEPE-TSDQTSTESATNNTNPE----- 628
Db 971 PPSPIHEKETEEFDS--HPESPVLSEKERERHQVTSSTTTTIVTREYNDEPQEQHQ 1028
QY 629 -----POQTATGSAHEETSTRDALQDQDSDSDVLPFGARYRAGPDRRSAR 682
Db 1029 GPHSPAPSSHTAEAPHIVETTTTITVTRFQEEPE-----LEYKQEDNSRKSPSSH 1081
QY 683 IOE 685
Db 1082 SOE 1084

RESULT 10
t29919
hypoetical protein ZC449.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: t29919
R:Latreille, P.; Gattung, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZC449.
A:Reference number: Z20708
A:Accession: t29919
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-550 <LAT>
A:Cross-references: EMBL:U41510; PIDN:AAA82634.1; CESP:ZC449.5
C:Genetics:
A:Gene: CESP:ZC449.5
A:Introns: 36/1; 73/3; 86/3; 110/3; 132/2; 469/3; 504/2

Query Match 4.4%; Score 199; DB 2; Length 550;
Best Local Similarity 19.3%; Pred. No. 0.0019;
Matches 102; Conservative 80; Mismatches 219; Indels 128; Gaps 16;

QY 249 LNKSCRTSLCYSEDGQELVSYSSDIYFLDPKDTARELK--TPSAERREELQPP 306
Db 116 LENKCLLTR-----ARRHAKTTPATDDEETTT 148
QY 307 VKRLRLGDSDTGPRAPSERERDEQSPNVSLMORMSDMLSRWFEEA-SEVAQSNRG 365
Db 149 VKSHKKKSTTDAKPTTEEPASTKKEKKTESKKA VKTTTAAPAESAETFPSEEN 208
QY 366 RGRSRPGGTQSQDISLTPVPSPDLEVSETAMEVDTPAEQFLQPTSTSTMSQAHS 425
Db 209 KTTSEESGNVSAEVTTSSTEVSEEQVKTSTAEATEEK-----STPASEEATTS 262
QY 426 SPTSPHSTPLSPDSEQRQSVASGHHTHQSDNNNEKLSPKPGTGEPVLSLHYSTEG 485
Db 263 SETSE-----TTERSEE-----ETTTKPA-----KTHKSKT 290
QY 486 TTTSTIKLFTDWSIAGSRGIGSHCKSEGOESFVQSSVQPPGDSQETKAPESSE 545
Db 291 TTTT-----EASTKDSK---SHKTKKEKVTTPEPTEESSESPTEEPQSSE 340
QY 546 DVTKYQGV--SAENPVENHINTQDKFTAKPLDLSNGERDNLDRSCGVPEESASSE 603
Db 341 STPTSEESTQSEKPD-----KKEDKKDKDKKKOKKED-----SEEDDDDK 384
QY 604 KAKEPETSQSTESATNNTNPEQFQTEATGSAHEETSTRDALQDQDSDSDVLP 663
Db 385 KSKSSSSDSKDEKSTSSDSSEASSEKTVGK-----KTLFSGPPDSDEDDDD--- 436
QY 664 IPGARYRAGPDRRSARAVARQIEFFRRRKERK---EMEELDTNIRPLVKMYKGRN 720
Db 437 -----EGAGADEAFFSEKQATEAPAELSRTTVAVSKKKMKKSGENFTF 482
QY 721 TWIKENFANFVMSGDCGHIFIDWRHTAHLMLLEADNHWVNCLOP 769
Db 483 LVI-----AGVFVGLVAGVYMFNKKERENLSTVDERELTIAP 524

RESULT 11

S38170
SRP40 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR092c; protein YKR412a
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38170; S40645; S37702
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38170
A:Molecule type: DNA
A:Residues: 1-406 <BAL>
A:Cross-references: EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40645
A:Molecule type: DNA
A:Residues: 1-406 <BOU>
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552
A:Experimental source: strain S288C
R:Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.
submitted to the EMBL Data Library, May 1993
A:Description: Interactions between three common subunits of yeast RNA polymerases I
A:Reference number: S37702
A:Accession: S37702
A:Molecule type: DNA
A:Residues: 1-399, 'N', 401-406 <LAL>
A:Cross-references: EMBL:L11275; NID:g295670; PID:g295671
C:Genetics:
A:Gene: SGD:SRP40
A:Cross-references: SGD:S0001800; MIPS:YKR092c
A:Map position: 11R

Query Match 4.4%; Score 198.5; DB 2; Length 406;
Best Local Similarity 23.2%; Pred. No. 0.0014;
Matches 89; Conservative 56; Mismatches 172; Indels 67; Gaps 9;

QY 386 VPSSPDLEVTAMEVDTPAEQFLQPTSTSTMSQAHSSTSPTEPHSTPLLS----- 438
Db 8 VDEVPKLSVKEKEEKKSS 67
QY 439 --SPDSQRQSVASGHHTHQSDNNNEKLSPKPGTGEPVLSLHYSTEGTTTS----- 489
Db 68 SDSSESSSESEDETK 127
QY 490 -----TIKLNFTDWSIIA--SSRGIGSHCKSEGOESFVQSSVQPPGDS 535
Db 128 KRARESDNEADAKETKKAKEPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 187
QY 536 ETKAPE-----SSDVTYQEGVSAENPVENHINTQDKFTAKPLDLSNGERDNLN 588
Db 188 ESDESDSQSSSDSD 245
QY 589 LDRSCGVPEESASSEKAKEPETSDQSTESATNNTNPEQFQTEATGSAHEET--ST 646
Db 246 ---SSGSSSD 302
QY 647 RDSALQDQDSDVLPFGARYRAGPDRRSARAVARQIEFFRRRKERKEMEELDTNIR 706
Db 303 NESTPSASSSSSSANKLNIPAGTDEIKGQRK-----HFSRVDRSKINFEAWE----- 349
QY 707 PLVKMYKGRNRTMIKEANFWG 730
Db 350 -LTDNTYKG-----AAGTWG 363

RESULT 12

T34513
hypoetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R:Accession: T34513
R:Favella, A.; Vaudin, M.
Submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: Strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 4.4% Score 198; DB 2; Length 3507;
Best Local Similarity 20.6%; Pred. No. 0.024;
Matches 160; Conservative 112; Mismatches 282; Indels 222; Gaps 33;
QY 77 ISNPYSKVLTTIRSGHRANIFSAKFLPCNDKQIVSCGDGVIFYTNVEODATNRQCC 136
DB 1994 ISSSTSKMTSSKSPENVTMSSPEVSTSSSKSTTASSTV---SSTPSSSSSEAP 2049
QY 137 FTCHYGTYEIMVNDPYFLSCGEGDGVTFTRIKTSCTKEDCKDDILLINCRRAATS 196
DB 2050 LTSSPATTTTEIT-----ESSVKSTTPKESSEIIVKLSSKPE 2089
QY 197 VAICPPPIYLAAGCSDSRVYDRRLMTRATNAGRGTTGMVAFIPSHLNANKSCRV 256
DB 2090 V-----TESSVK-----SSPTSTTSQSVTSTVPETSKSTVLSSEAPV 2128
QY 257 TSL-----CYSEGDQELVSYSDYIYLPDKDTARELTPPAERRE----- 300
DB 2129 TSTPTTEVHTSETKPSLSASSTGDTNSTTPTSSSLASVKSTSAPEGTSASVAVKLSS 2188
QY 301 ---ELRQPPVKRL-----RLRGDWSDTGPRARPESERE-----RDGQSPNVSLMQ 343
DB 2189 LSPDVQSSTKTFDASTVQASSTSGTSVKSTSEPSHVTKLSTSSNPSSVPVTS 2248
QY 344 RMSDMLSRWFEEASEVAQSNRGRS-RPRGTSQ-----SDISTLPTVP 387
DB 2249 PKS---TPTVPESTEQPTSTTPSGQSLAPMNSSEVLATSEPHVLSLSLSPDVQSSTTP 2305
QY 388 S-----SPDLEYSETAMEVD-----TPAQFLOP-----STSSSTMSQAHS--- 423
DB 2306 NNLSESTVETPKTSSEVSLNSEESTTEAPTLLSPDILSTTNLSQSSYSTVEDRSI 2365
QY 424 TSSPTSEPHSTPLL-----SSPD---SPQSVASGHHHTHQSDNNNEKL---SP 468
DB 2366 SSENSEKPTSAPELYTSSVTHVASSPDVPTSESPDLDLTSSTENIPASSKQITSTP 2425
QY 469 KPGT---GEPVLSLHYSTEGTTSTI---KLNFTDWS--IASSRGIGSHCKSEGE 519
DB 2426 TPDUTTASEPTKTSMSPDLTSTTNVLSSESTTPSSSKSPVSSSTEGISVVTST---E 2482
QY 520 ESFVPOSSVQ-----PPEGDSKKAPEESEDVTK-----YQGVSAEN- 558
DB 2483 FKVPSTISSVLEEDLTITTPSPILEETITASETSEPTEDSLTVSVRIHELTTSEN 2542
QY 559 PVENHINITQSKFTAKPLDSNGERNDLNRSCGVPESASSEKAKEPE----- 609
DB 2543 PRESESTTTSSSE--SSKPSQEPAG-----ILTSTVVVPTSSVSLITASEIATSTNTPK 2595
QY 610 -----TSDQSTESATNENN-TNPEQFQTEATGPSA-----HEETSTRDSAL--- 651
DB 2596 QGRTPTTPKSLVKSTTSPTSSPESTKRTVSTVSTVTTTTEETTESILTA 2655
QY 652 -----QDTRDSDDDVPLIPGARYAGPDRRSAAVARIQEFFRRRKERKEMEELDT 701

DB 2656 APSKPTSTESSEAPT-TPAKYSETKPSNVNST-----SRKSTENVET 2698
RESULT 13
T06310
hypothetical protein F11C18.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
R:Accession: T06310
R:Revan, M.; Terry, N.; Ardiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
Submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15589
A:Accession: T06310
A:Molecule type: DNA
A:Residues: 1-852 <BEV>
A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:F11C18.80
A:Experimental source: cultivar Columbia; BAC clone F11C18
C:Genetics:
A:Gene: ATSP:F11C18.80
A:Map position: 4
A:Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3
Query Match 4.3% Score 193; DB 2; Length 852;
Best Local Similarity 21.9%; Pred. No. 0.0071;
Matches 129; Conservative 76; Mismatches 224; Indels 160; Gaps 26;
QY 162 EDGVRWFDTRIKTSCTKEDCKDDILLINCRRAATVAICPPPIYLAAGCSDSRVYDR 221
DB 365 ESSSVKQADLS-KSDDIKEETEPALLDSKDLTS---PPV-----DSSV----- 405
QY 222 RMLGTRATNAGRGTTGMVAFIPSHLNANKSCRVTSCLYSEGDQELVSYSDYIYLPD 281
DB 406 ---TAATSS-----ENKNSV-----QILPKSTSG----- 428
QY 282 PKDITARELKTPSAERREELRPQPKRLRLKGDWSDTGPRAPESEEREDEQSPNVSL 341
DB 429 --DETA-NVSSPSMAELPEQSV-----KKTANOKKESSTE---EVKPSASI 471
QY 342 MORMSDMLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLPTVP-----SSPDLEV--- 394
DB 472 AT-----EEVSEFNTSEPO-VTKKSKKVASSSKTKPTVPPSKKSTSETKVAKQ 520
QY 395 -----SETAMEVDTPAEQFLQSTSTMSQA-HSTSSPTESP--HSTPLLSPDSQ 444
DB 521 SEKKVVGSDNAQESTKPEKKKPGRAIDESLHTSGDNEKPAVSSGKLASKKKEA 580
QY 445 ROSVEASGHHHTHQSDNNNEKLSPKPGTGPVLSLHYSTEGTTTSTIKLNFDEWSSIAS 504
DB 581 KQTV-----ESPNSNTKRLSLGQGA-----SGESLVGSRKIV---WPMDOA 622
QY 505 SSRGIGSHCKS-----EGQESFVQSVQSPGDSKTPAES-----EDVT 548
DB 623 YKGVVSTDAKAKHLVIYDGDQELLYLKNQKSPDSELSQDEEADQTEGDEAS 682
QY 549 KYQEGVSAAE---NPVENHINITQSKFTAKPLDSNGERNDLNRSCGVPESASSEK 604
DB 683 TVGSGAGSSKAKATPASKSKTSQDDKTASKSKDSKEASR-----EEEASSEE 730
QY 605 AKPEPESDQ---STESATNENNTNPEQFQTEATGPSAHEETSTRDSALQDTRDSDDD 661
DB 731 ESEEEPPRTVKGSGSRSKKDISSVSKSGKSKASSKKKEEPSKATSS-----KSKSGP 785
QY 662 VLIPIGARYAGPDRRSAAVARIQEFFRRRKERKEMEELDTLNIRRLPVK 710
DB 786 VKSVPAKSKTGKAKSGSASTPA-SKAKESASESESEETPKPEPATK 833
RESULT 14
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.13) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucanase; mucin-like protein MUC1; protein YIR01
C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
 C:Accession: S48478; A26877; B26877; S27281; JC6123
 R:Rowley, K.
 Submitted to the EMBL Data Library, October 1994
 A:Reference number: S48478
 A:Accession: S48478
 A:Molecule type: DNA
 A:Residues: 1-1367 <ROW>
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GSPDB:GN00009; MIM
 R:Yamashita, I.; Nakamura, M.; Fukui, S.
 J. Bacteriol. 169, 2142-2149, 1987
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STAI.
 A:Reference number: A91831; MUID:87194600; PMID:3106330
 A:Accession: A26877
 A:Molecule type: DNA
 A:Residues: 1-242 <YAM>
 A:Cross-references: EMBL:M16164; NID:G172522; PIDN:AAA35014.1; PID:G172525
 A:Accession: B26877
 A:Molecule type: DNA
 A:Residues: 762-1331 <YA2>
 A:Cross-references: EMBL:M16165; NID:G172523; PIDN:AAA35015.1; PID:G172526
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
 FEBS Lett. 239, 179-184, 1988
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
 A:Reference number: S27281; MUID:89031230; PMID:3141213
 A:Accession: S27281
 A:Molecule type: DNA
 A:Residues: 1-31 <PAR>
 A:Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:G4552
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
 A:Title: Muc1, a mucin-like protein that is regulated by Msl0, is critical for pseudohy
 A:Reference number: JC6123; MUID:9632237; PMID:8710886
 A:Accession: JC6123
 A>Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1367 <LAM>
 A:Cross-references: GB:U30626; NID:G1304386; PIDN:AAC49609.1; PID:G1304387
 C:Genetics:
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
 A:Cross-references: MIPS:YIK019c; SGD:S0001458
 A:Map position: 9R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
 F:5-21/Domain: transmembrane #status predicted <TM>
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 4.2%; Score 190.5; DB 1; Length 1367;
 Best Local Similarity 22.3%; Pred. No. 0.018;
 Matches 124; Conservative 64; Mismatches 270; Indels 99; Gaps 18;

Db 151 PNDPTFLSCGEGCTVRFDTRIKTSCTKEDCKDDILNCRRATSAICPPPIYLAVG 210
 Db 272 PTPPTTSCVKEKPPPHDFT---TPCTKK--KVTSTKCTKKT---PVPTPSSST 321
 QY 211 CSPSSVRIYDRRLMTRATNGYAGRTGMVARFIP-----SHLNKSCRVTSLCYSEDQ 266
 Db 322 TESSAPVPTPSSSTTESSAPVPTSSSTESSAPVPTPSSSTTESSAPVPTSSSTESSA 381
 QY 267 EILVSYSSDIYFLDPKDDTARELKTSAERRELRQPPVKRLRLRGDWDSDTGPRARPE 326
 Db 382 PVTSTT-----ESSAPVPTPSS-----STTESSAPV 410
 QY 327 SERERCEQSPNVLQMRMSDMLSRWFEAEAEVAQSNRGRSRP-RGCTQSQSDISTLPT 385
 Db 411 TSSTTESSAP-----VTSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 460
 QY 386 VPSSPDLEVE---TAMEVDTPAEQFLQPTSTSTMSAQAHSTSPTES---PHSTPLLS 439
 Db 461 -PSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTPSSST 519
 QY 440 PDSEQRQSVASGHHHTHQ-----DNNEKLS---PKPGTGEPLVLSHYSTGTTSTIK 492

Db 520 TESSAPAPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSSTPVTSTTESSA 579
 QY 493 LNFIDENSIASSRIGSHCKSEGESEF---VPOSSVQPPGEGDSETKAPESESDVTK 549
 Db 580 PVPTPSSSTTESSAPVPTPSSSTTESSAPAPTSSSTTESSAPVPTSSSTTESSAPVP 639
 QY 550 YQEGVSAEN---PVENHINITQSKFTAKPLDSNGERNDLNDLRSCGVPPEASAKAK 606
 Db 640 TPSSSTTESSAPVPTPSSSTTESSAPVPTPSS-----TTESSAPVPTSSSTESS 692
 QY 607 EPETSDQFSTESA-----TNENNTNPEQFQTEAT-----GPSAHEETSTRDSA 650
 Db 693 APVTSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 752
 QY 651 LQDTPDDDDPVLPGA 667
 Db 753 TSSSTESSAPVPTPSS 769

RESULT 15
 A41859
 Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
 C:Species: Haemophilus influenzae
 A:Variety: strain HK715
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
 C:Accession: A41859
 R:Poulsen, K.; Reinholdt, J.; Killian, M.
 J. Bacteriol. 174, 2913-2921, 1992
 A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae
 A:Reference number: A41859; MUID:92234949; PMID:1373717
 A:Accession: A41859
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1702 <POD>
 A:Cross-references: GB:M87489; NID:G148906; PIDN:AAA24966.1; PID:G148907
 A:Experimental source: strain HK715
 A:Note: Sequence extracted from NCBI backbone (NCBIP:97282)
 C:Superfamily: Iga-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

Query Match 4.2%; Score 188; DB 2; Length 1702;
 Best Local Similarity 18.9%; Pred. No. 0.032;
 Matches 150; Conservative 120; Mismatches 311; Indels 212; Gaps 33;

QY 67 LSGSDDTKLVISNPYSRKVLTIIRSGHRANIFSAKFLPCTNDKQIVSCGDGVITYTWE 126
 Db 773 ITASDNAKVHIG--YKAGDTVCVRSYTG-----YVCTTDK--LSDKALNSFNATNVS 822
 QY 127 QDAETNRQOFTCH---YGT---TYEIMTVPNDPYTFILSCGEGDGTVRWFDTRIKTSC 177
 Db 823 GNYNLSGNANFVLGKANLFGTISGTGNSQVRLTENSHTWLTG---DSNVNQLNLD----- 874
 QY 178 TKEDCKDDILNCRRATSAICPPIP-----YLLAVGSDSVRIYDRRLMTR 227
 Db 875 -----KGTHLNAQNDANKVTTNTLTVNSLSGNGSPYLL---TDLNKGQDKVVVTKS 925
 QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSCRVTSLCYSED---GOETLVSYSSDIY 277
 Db 926 ATGNFTLQVADKTGETPKNELTLFDASNATNNLNLVGNVTLGAWKYKLRNNGRY- 984
 QY 278 YLFDPK-----DDTAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318
 Db 985 DLYNPEVEKRNQVDTNTITPNNIQADVPSVPSNNIEARVETPVPPAPATPSETTET 1044
 QY 319 TGPRAPES-----ERERDG-----EQSPNVLQMRMSDMLSRWFE----- 354
 Db 1045 VAENSKQESKTEVKNQDQDATTETAQNGEVAEAKPSVKANTQTNEVAOSGETEETQTE 1104
 QY 355 --EASEVAQSNRGRSRPRRGTSQSDISTLPTVPS-----SPDLEVESETA 398
 Db 1105 IKETAKVEERAKVEEKAKVEKDEIQEAPQWASETSPKQAKPAKVEYSTDTKVETQ 1164
 QY 399 MEVDTPAEQFLQPTSTSTMSAQAHSTSPTESPHSTPLLSPPDSEQRQSVASGHHHTHQ 458

Db 1165 VQA-----QFQSTQTVAAEAAT-----SPNSKPAETOPSEKTNAPETVPVVKNO 1211
QY 459 SDNNNEKLSPKPTGEPVLSLHYSTGTTSTIKLFTDEWSSIASSSRGIGSHCKSEGG 518
Db 1212 TENT-----TDQP-----TEREKTAKEKTOEPPOVASQA-----SPKQEQ 1249
QY 519 ESFVPOSSVQPPGSEKAPESSEDTKYQEGVSAENPV--ENHIN-----ITQSDK 571
Db 1250 SETVQPOAVLESENVTNNAAEEVQAOLOTOATVSTKOPAPENSINTGSAITATETAE 1309
QY 572 FTAKPILDSNGSERNLNLDRSCGVPESSASSEKAKPETSDQSTESATNENNTNPEPQF 631
Db 1310 KSDKP-----QTAASTEDASQKANTVADNSVANNSESSEPKSR 1351
QY 632 QTRATGPSAHEETSTROSALQDTHDDDDVLPYGARYRAGPDRSASAVARIQEFFERRK 691
Db 1352 RRSISQP---QETSABETTAATDET-----TADNSKKRSP--NRRS-----RRS 1392
QY 692 ERKEMEELDTNIRRLPVLMVKYGHRSRTMIKEANFWANFVMSGSDCGHIFI---WDR 748
Db 1393 VRSE-----PVTNGSDRSTVALRDLTSTNTNAVISDAMAKAQFVALNVGR 1438
QY 749 HTAEHMLMLLEADN 761
Db 1439 AVSQHISQLEMN 1451

RESULT 16
F95133
Immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95133
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Hickey, E.K.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75263.1; PID:g14972632; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp154
C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 4.1%; Score 187.5; DB 2; Length 2004;
Best Local Similarity 18.2%; Pred. No. 0.042;
Matches 121; Conservative 99; Mismatches 237; Indels 209; Gaps 23;

QY 111 IVSCSGDVIFTYTNEQDAETNROCQFCHVGTVEI---MTVPND--PYTELSCGSD 163
Db 139 LVAGTNGVIVSVHALEN-----HLLNTNTDYELTSGEKLPLPREISGTYTGYIKE 191
QY 164 G-----TVRFDTTRIKTSCTK----- 179
Db 192 GKTTSESEVSNQKSVAPFTKQKVYDYNVTFNFDHPSTVQAIQEQTPVSTKTPVQVW 251
QY 180 ECKDDIILNCRRAATVAICPPPIPYLAVCCSSSVSRIYDRMLGTATNGYAGRTWG 239
Db 252 EKPFSTELINPRKEKQ-----SSDSQEQLAEHKNLETKKEEKISPKETG 297
QY 240 MVAREFIPHLNKKSCRTSLCYSEDGOEILVSYSS----- 274
Db 298 VNTLNPQDEVLISGQLNKKELLYRETMETKIDFQOEIQENPDLAGTVRVKQEGKLGKV 357
QY 275 DYIYLFDP-KDDTARELKTSPSAERREELRQPPVKRLRLRGWDSGTGPRAPESREDG 333

Db 358 EIVRFSVNKEEVSREIVSTST-----TAPSPRIVEKGTKKTKQVIKEQPGTVEHKD 409
QY 334 EQS-----PNVSLMQRMDSMLSRWFEEASVQAOSNRGRSRPRGGTTSQSDISTLPTV-- 386
Db 410 VQSGAIVEPAIQ-----PELPEAVVSDKGEPEVQP-----TLPEAVV 446
QY 387 -----PSSPDLEVSETA-----MEVDTPAEQFLQPSSTSSMSAQ 420
Db 447 TDKGTEVQEPESDPTVVSDKGEPEQVAPLPEYKGNIEQVAPETPVKTKBQGPKEKTEEV- 505
QY 421 AHTSPTSPESPHSTPLSSPDSEORQSVASGHHTH--HOSDNNNEKLSPKPGTGEVLS 478
Db 506 -----PVKTEETPNPNEGTEGTSIOEANPVQPAEESTTNSEKVP----- 549
QY 479 LHYSTEGTTTSTIKLFTDEWSSIASSSRGIGSHCKSEGESEFVPOSSVQPPGSEK 538
Db 550 ---DTSSKNTGEVSSNPDSSTTSVGSNKPEHNDKSNSEKT--VEEVPYNNEGTVEGT 605
QY 539 APESSEDTKYQEGVSAENPVENHINITQSKFTAKPLDSNGSERNLNLDRSCGVPEE 598
Db 606 SNOETKPVQPAEE-----TQTN--SGKIANENTGEVSNKPSDSKPPV-EE 648
QY 599 SASSEK---AKEPETSQDSTESATNENNTNPEQFQTEATGPSAHEETSTROSALQDTH 655
Db 649 SNOPEKNGTATKPENSGNTTSEN---GQTEPEPSNGNSTEDVSTESNTSNGNEEIKQ 704
QY 656 DSDDDP 661
Db 705 ENELDP 710

RESULT 17
H64106
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (str
N;Alternate names: Immunoglobulin A1 proteinase type 1
C:Species: Haemophilus influenzae
A:Variety: strain Rd KW20
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
A:Accession: H64106; A41500
R:Geynechmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64106
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1694 <TIGR>
A:Cross-references: GB:U32779; GB:I42023; NID:g1574009; PIDN:AAK22651.1; PID:g1574019
A:Experimental source: strain Rd KW20
R:Grund, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A:Title: Localization of the cleavage site specificity determinant of Haemophilus inf
A:Reference number: A41500; MUID:90129281; PMID:2105270
A:Accession: A41500
A:Molecule type: DNA
A:Residues: 1-377 <GRU>
A:Cross-references: GB:X59800
A:Experimental source: strain Rd KW20
A:Note: the authors translated the codon TGG for residue 319 as Thr
C:Function:
A:Description: this proteinase is classified as type 1 because it cleaves at a prolin
C:Superfamily: IgA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 4.1%; Score 186; DB 2; Length 1694;
Best Local Similarity 19.4%; Pred. No. 0.041;
Matches 153; Conservative 113; Mismatches 310; Indels 214; Gaps 33;

QY 67 LSGSDDTKLIVSNPYSRKVLTIIRSGHRANIPSAFLPCTNDKQIVSCSGDVIFTYTNE 126

Db 773 ITASNAKVHIG--YKAGDTVCVRSDYTG-----YVCTTDDK--LSDKALNSFNATNVS 822
QY 127 QDAETNRQCFCH-----YGT-----TYEIMTVPMNDPTFLSCGDGTVRFWFDTRIKTSC 177
Db 823 GNVALSGNANFVLKCANLFGTISGTGNQVRLTENSWHHLTG-----DSNVQNLND----- 874
QY 178 TKEDCKDILLNCRAATSVAICPPIP-----YLLAVGSDSSVRIYDRMLGTR 227
Db 875 -----KGHIHLNAONDANKVVTYNTLTVNSLSGNGSFYLL-----TDLNKGQDKVVVTKS 925
QY 228 ATGNV-----AGRGTTGMVAREIPSHLNKSKRVTSCLCYSED-----GOETLVSYSSDI 277
Db 926 ATGNFTLOVADKTEPTKNELTFLDASNATRNLLNSLVGNVTLGAWKYKLRNVNGRY- 984
QY 278 YLDPDK-----DDTAR-----ELKTPSAERRELEQ-----PPVKRLRLRGDMSD 318
Db 985 DLYNPEVEKRNQVDDTNTITPNNIOADVPSPVSNNEIARVETPVPPAPATPSETTET 1044
QY 319 TGPRAPESERERDEQSPNVSLMORMDMLSRWFEA-----GEVAEAKPSVKANTQTNEVAQSGSETEE 1099
Db 1045 VAENSKQESKTVEKNEQDATETTAQN-----GEVAEAKPSVKANTQTNEVAQSGSETEE 1099
QY 363 -----NRGRSRPRGGTQSIDSITLPTVPS-----SPDLEVSSETAMEV 401
Db 1100 TQTEIKETAKVEKEAKYKDEIQAPQOMASSETSPKQAKPAPKEVSTDTKVEETQVQA 1159
QY 402 DTPAEQFLOPSTSTMSQAQAHSTSPPTSPHSTPLSSPSEQRSQVSEAGHHTHQSDN 461
Db 1160 -----QPOTOSTVVAEAT-----SPNSKPAEETQPSKNTAEPVTPVVSKNOTEN 1206
QY 462 NNEKSPKPGTGPVLSHYSTEGTTSTIKLNFDTDEWSSIASRRGIGSHCKSEGOEES 521
Db 1207 T-----TDQP-----TEREKTAKEVTEKTPQPPQVQA-----SPKQESQET 1244
QY 522 FVQSSVOPPEGSETKAPESEEDVTKQEGVSAENPV-ENHIN-----ITQSDKFTA 574
Db 1245 VQPOAVLESENPTVNAEVOAQLOTQTSATVSTKOPAPENSINTGSATAITAEKSD 1304
QY 575 KPLDNGSRNDLNLDRSCGVPPEASSEKAKEPETSDQTSATNENNTNPEPQFQTE 634
Db 1305 KP-----QTETRASTEDASQHKANTVADNSVANNSESSOPKSRRRS 1346
QY 635 ATGSAHEETSTDSALQTDSDDDPVLIPGARVRAGPGRDRSARVARIQEFRRRKRK 694
Db 1347 ISQP-----QETSAEETAASTDET-----TIADNSKRKP-NRRS-----RRSVRS 1387
QY 695 EMEELDTLIRPLVMYKGRNKRMTIKNEANFGANFVMSGDCGHIFI---WDRHTA 751
Db 1388 E-----PTVNGSDRSTVALRDLTSTNTNAVISDAMAKAQFVALNKGAVS 1433
QY 752 EHLMLLEADN 761
Db 1434 QHISOLEMNN 1443
RESULT 18
T29612
hypothetical protein D1014.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29612
R:Du, Z.; Leimbac, D.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid D1014.
A:Reference number: Z20652
A:Accession: T29612
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-727 <DUZ>
A:Cross-references: EMBL:U53180; PIDN:AAA96284.1; GSPDB:GN00023; CESP:D1014.5
A:Experimental source: strain Bristol N2; clone D1014
C:Genetics:
A:Gene: CESP:D1014.5

A:Map position: 5
A:Introns: 52/1; 379/2; 442/2; 579/3; 618/2; 647/3; 676/2; 707/3
Query Match 4.0%; Score 183; DB 2; Length 727;
Best Local Similarity 23.5%; Pred. No. 0.02; Indels 88; Gaps 14;
Matches 89; Conservative 47; Mismatches 155; Indels 132; Gaps 28;
QY 332 DGEQSPNVSLMORMDMLSRWFEASEVAQSNRGRGRPRG-----GTQSQDSTLPTVP 387
Db 43 DSKASPVSDRQOYSD-----QKQEVVNSNNTESKNAGSENPSSSETADQATNEP 95
QY 388 SSPDLEVSSETAMEVDTPAEQFLOPSTSTMSQAQAHSTSPPTSPHSTPLSSPSEQRSQ 447
Db 96 -KPKVSDDNAGSAETIESNGIEKIPS-----NDTSPFDSAVVP--SGPAQEEDEA 145
QY 448 VE-----ASGHHTHQSDNNNEKLSKPGTGPVLSLH-----YSTEGTT----- 487
Db 146 VEEKKVEEHEENEGNSDAEENKSGDAVTTITELSMEDREGNDPKDESGTIVTSNE 205
QY 488 -TSTIKLNFDTDEWSSIASSS---RGIGSHCKSEGOEES-----FVP 524
Db 206 ETTTTEKNKVADEAPSKAEISQASQSDSHVKEGESETTTTSDSATTEKKTNGAELVE 265
QY 525 QSSVQPPGEGSETKAPE---ESSEDVTKYQGVSAENPVENHINTQSDKFTAKPLDSS 581
Db 266 KSETEKNGDSSTKSPVETSESSVDTTIEQLASNDDEE-----KQVDGNE 312
QY 582 GERNDLNLDRSCGVPPEASSEKAKEPETSDQTSATNENNTNPEPQFQTEATGPSAH 641
Db 313 KESTFRVLNEDHNEENEEIGETK---KVTEETTETTERQNVDDTENTQTTTSSSY 368
QY 642 -----EETSTRDSALQDQTD 656
Db 369 NAKEEKSYRKCKVEWND 387
RESULT 19
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F90073
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F90073
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2271 <KUR>
A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA43752.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2447
Query Match 4.0%; Score 183; DB 2; Length 2271;
Best Local Similarity 21.1%; Pred. No. 0.087;
Matches 147; Conservative 101; Mismatches 316; Indels 132; Gaps 28;
QY 55 NTICWNTDGEYLSCGSDDTKLVISNPYRKVLTTTR-----SGHRANIFSAKFL 103
Db 939 NSIANSQSASTSKSDSQSTSLSTSDSKMSSTSESLDSTSTSGSVGSLIAASQSVS 998
QY 104 PCTND-----KQIVSCSGDGVIFYVTNVEQDAETNRQCF-----TCHYGTTEIM-----T 149
Db 999 TSTSDSMSTSEIVSDS-----ISTSGLSASDSKMSVSSMSSTSGSQSTSESLDSQST 1053
QY 150 VPNDPYTF-LSCGEDGTVRWFDTRIKTCTCKEDCKDDILNCRRAATSAVACPPIPYLA 208
Db 1054 SDSDSKSLSLSTSQSGST---STSTSTASVRTSESQSTSGSMSASQSDSMS-----IS 1104

209 VCSOSSVRIYDRMLGTRATGNVAGRTTGMVARFPTPSHLNNKSCRVTSLCYSDGOEI 268
|||: : : | :|: : :
1105 TSFSDTSDSKASTASSTISQSASTSTSGVS-----TSTSLSTNSER 1150
|||: : : | :|: : :
269 LVYSYSDYILFDPKDDTARELXTPGAEEERELRPPVKRLRLRGDWSDTGPRARPESE 328
|||: : : | :|: : :
1151 TSTSVSDTSLTSTSESDISE-STSTSDSISEALSA-----SESTSISESBN 1197
|||: : : | :|: : :
329 RERDGE-QSPNVSLMORMDMLRWFEAEVAQSNRGRBS-----RPRGTSOS---- 378
|||: : : | :|: : :
1198 STSDSESQASAFLESLSSEST---ESTSESVSSSTSESTSLSDSTSESGSTSTSLNS 1254
|||: : : | :|: : :
379 -----DISTLPTVPSPDLVESE-----TAMEVDTPAEQFLQPTQSTSTMSAQAH 422
|||: : : | :|: : :
1255 TSGSASISTSTISESTSTFKSESVSTSLSMSTSTSLNSSTSLSTSLSDSTSDSKS-DSL 1313
|||: : : | :|: : :
423 STSSPTSEPHUSTPLLLSPDSEQRQSVASGHTTHQSDNNNEKLSPKPCTGEPVLSLHYS 482
|||: : : | :|: : :
1314 STSMSTSDSIST---SKSDSIS-TSTSLGSTSESDSTSSSESKSDSTSMI-SMSOS 1368
|||: : : | :|: : :
483 TEGTTTSTIKLNFTEWS-----SIASSRGIGSHCKSEGOEESFVPOSSVOPPEGDET 537
|||: : : | :|: : :
1369 TSGSTSTSTSLSDSTSTSLSLASMNQSGVDSNASQSASNS---TSTSTSESDSOS 1424
|||: : : | :|: : :
538 KAPEESEDYTKQEGVSAENPVENHINITOS-----DKETAKPLDMSGERNLMLDRS 592
|||: : : | :|: : :
1425 TSYTYSOS--TSQSESTSTSLSDSTSIKSTSGSGSTSTSLASLGSSESDSQSISTS 1482
|||: : : | :|: : :
593 CG-VPEESASSEKAKEPETSQDTSTESATNNTNNTNPEQFQTEAT----GPSAHEFTSTR 647
|||: : : | :|: : :
1483 ASBSTSEASATSLSDSTSTSGSGASTSTSLNSASASESDSDSTSLSDSTSASMOSSES 1542
|||: : : | :|: : :
648 DSALQDITDDSDDDPVLIPGARYAGPGRRRSAVARI 683
|||: : : | :|: : :
1543 DS--QSTSLASLSDSL-----STSTSNEMSTIASL 1569
|||: : : | :|: : :
RESULT 20
F89500
protein T27A10.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C:Accession: F89500
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID: 99069613; PMID: 9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F89500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <STO>
A:Cross-references: GB:chr_X; PIDN:AAB52504.1; PID:g1326288; GSPDB:GN00028; CESP:T27A10.
C:Genetics:
A:Gene: T27A10.6
A:Map position: X

Matches	131;	Conservative	79;	Mismatches	258;	Indels	184;	Gaps	20
Qy	295	AERREELRPPVKRLRLGDWSDTGR----	ARPESEREDGEQSPNSILMQRMSDL	349					
		: :	: :						
Db	246	AERLQRLOEQEAYLLSLQHDHRRPHSOQPPPOOERSKPFHAEPKAHYEPAD--	303						
Qy	350	SRWFEEASVAQNRRGRRPRGGTSQSIDISTLTTPVS-----	388						
Db	304	----RAREVEDRPRTKNHSPEAQSTQTGYLEPVPSPRSSEFNGNSEVHPALQRPRA	358						
Qy	389	SPDLEVSETAM-----EVDTPAEOLFQ-----	PSTSTWSAAQHSTSS	426					
		: : :	: : :						
Db	359	EPQVQNSHLASLKNNVSPVSRSHSFSDPSPKFAHHLLRSQDPCPPSRSEVLSSQSDSKSE	418						
Qy	427	-----PTESPHTPTLLSSPDSPQRSVEASGHHTTHHOSDNHN--	463						

RESULT 21

T46481 hypothetical protein DKFZp434A025.1 - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C.Accession: T46481
A:R.Duisterhoft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46481
A:Status: preliminary
A:Molecule type: mRNA
A:Studies: 1-1027 <AA>
A:Cross-references: EMBL:AL137755
A:Experimental source: adult testis; clone DKFZp434A025
C:Genetics:
A:Note: DKFZp434A025.1

Query Match 3.9%; Score 178; DB 2; Length 1027;
Best Local Similarity 20.1%; Pred. No. 0.058;
Matches 131; Conservative 79; Mismatches 258; Indels 184; Gaps 26;

[illegible]

Db 419 APDPTOKANSRSDSDVPPRPVPTTSRSPVLSRRDS-----PLOGSGQQNSQAGQRNSTS 474
QY 464 -----EKLSPKPGTGEVPLSLHYSTEGTITSTIKLNFTEWSSSIASSSRGIG-- 510
Db 475 SIEPRLLWERVEKLVPKPGSG-----SSGSS-----NSGSQPGSHPGSQSGGER 520
QY 511 -----SHCKSEGOEESFVPOSSVQPPGDSSETKAP--EESSEDVTKYQOEGVSAENPVENHI 564
Db 521 FVRSSSKSEGPSQRL--ENAVKEDKKEVFRPLKPAGEVDLTALAKELRAVEDVRPPH 579
QY 565 NITQSDKFTAKPLDSSNGSRNDLNL-----DRSCGVPER--SASSEKAKEPEPESDQTSFE 617
Db 580 KVTD--YSSSEESGTTDEEDDDVEQEGADSTSPETRAASSLNLNGETE---SVK 633
QY 618 SATNENNTPEQFQTEARGPSAHEFTSTRDSALQDQDSD-----DPVLI----- 664
Db 634 TMIVHDDVESEFAMTPSKEGTLIVRQTSASSTLQKHSSSFTPTIDPRLLQISPSGT 693
QY 665 -----PGARYRAGPD--RRSAVARIQEFFRRKERMELDLTINRRPLVKV 712
Db 694 TVTSVVGFCSDMRPEAIRQDTRKGVSVNVNPTNTR-----PQSDTPEIRK----- 740
QY 713 YKGRNSRPTMIKEANFWGANFYMGSDCGHIFIDRHTAHLMLLE--ADNHVNCLOPH 770
Db 741 YKKRFNSEILC--AALGWVNLIV-GTESG-----LMLLDRSCQGVYPLINRR 785
QY 771 PRDPLASSGIDYDIKINSPLEESRIFNRKLADEVITNELMLETNTITV 822
Db 786 RFOQMDVLEGLNVLTISGKKDKLVYLSWLNRKILHNDPEVKKQGWTTV 837
RESULT 22
T48933
WD repeat domain protein - Arabidopsis thaliana
N:Alternate names: protein F14L2.80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T48933
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225008
A:Accession: T48933
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1051 <OR>
A:Cross-references: EMBL:AL353818; GSPDB:GN00061; ATSP:F14L2.80
A:Experimental source: cultivar Columbia; BAC clone F14L2
C:Genetics:
A:Gene: ATSP:F14L2.80
A:Map position: 3
A:Introns: 36/3; 121/3; 205/3; 336/2; 668/3; 739/1; 773/1; 853/2; 894/2; 932/3
Query Match 3.9%; Score 178; DB 2; Length 1051;
Best Local Similarity 18.8%; Pred. No. 0.059;
Matches 128; Conservative 76; Mismatches 212; Indels 264; Gaps 26;
QY 42 KLEATLVHVDGCVNTICWNTDGEYILSGSDTKLVI--SNPYSR-----K 84
Db 50 RLLATLRHFGVNCVNRWAKNSRYVSGSDQVQIHERPFGSGTTEFGSGEAPDVENWK 109
QY 85 VLTIRSGHRANFSAKFLPCTNDKQIVSGSGGVPIYTNV----- 125
Db 110 AVMTLR-GHTADVVDLWNSP--DDSLASGLDNTVHNNRTGMCITVLRLHLSLVKG 166
QY 126 -----EODAEF-----NQCOFTC-HYGTTY 145
Db 167 TWDPIGSFASQSDDKVIVIRTSNDGMAHRTDGHNAKSLGSLGSGSPGCHFLTIT 226
QY 146 EIMTVPN-----PYTFL-----SCGEDGVVRWFDR 172
Db 227 HGQPKPKHSAVPLRGESVAYDFLGHSAPIIVVRNHSFKRIPSTHETKQVW---- 282

QY 173 IKTSCTKEDCKDDLLINCRRAATSVAlCPIPPYPLAVGCSDDSVRIYDRMLGTRATGNY 232
Db 283 -SNGTSKSGEKDLQSYN-----VIAMGSDQRTITVW----- 312
QY 233 AGRTTGMVAR--FIPSHLNKSKRVTSCLCYSDGQGLVSYSSDYLYL--FDPKDDTAR 288
Db 313 -----TTGS--ARPLFAKHFQGS--VVDLSWSPDGYSLFACSLDGTVMATHFDPK----- 360
QY 289 ELATPSAEERERELRQPPVRLRLRGDWSOTGPRARPESERERKDEQSPVNSLMORMSDM 348
Db 361 ELGVRLTDELDELKKSRYGDVR-----GQANLVESPAQLLLE----- 399
QY 349 LSRWFEAESEVAQSNRGRSRPRGGTQSQSDISTLPVPPSPDLEVSETAMEVDTPAEQF 408
Db 400 -----FASTQAGSKRAASDVQOQVTKKPSV-----SVETAKRRKSQVDDR 442
QY 409 LQSTSTSWAQAHSTSSPTESPHSTPLSSPDSEQSQSVASGHHHTHQSDNNNEKLS 468
Db 443 NKAESTGOTLNKASTLNRYSPVQKYYRRPDRKRRIPEAVG---VPOENN----- 493
QY 469 KPCTGEPVLSLHYSTEGTITSTIKLNFTEWSSSIASRGIGSHCKSEGOEESFVPOSSV 528
Db 494 -----IMIN-----GESHFLPASAA 509
QY 529 OPPEGDSKTAPEESSEDVTKYQGVSAENP-----VENHINITQSDKFTAKPLDSNS 581
Db 510 APAKGDGDPVEISNRDLGKE--IVCRNPDLKRSRTARATITTES--LVIEKPGTS 565
QY 582 GERNDLNLDRSCGVPEESAS 601
Db 566 GRDGLNVQSVGIKESST 585
RESULT 23
D86731
hypothetical protein yihD [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D86731
R:Boilotin, A.; Winker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
submitted to the Protein Sequence Database, April 2000
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: D86731
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1063 <STO>
A:Cross-references: GB:AE005176; PID:g12723779; PIDN:AAK04950.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yihD
Query Match 3.9%; Score 177.5; DB 2; Length 1063;
Best Local Similarity 24.3%; Pred. No. 0.064;
Matches 84; Conservative 46; Mismatches 163; Indels 53; Gaps 8;
QY 317 SDTGPRARPERERDGEQSPVNSLMORMSDWLSRWFEEAESEVAQSNRGRSRPRGGTS 376
Db 613 SQSSSSASSSSAAASSTSSSNVS-----SNTSSNSSEANTSSSTSNASSSSSSSEGSAA 667
QY 377 QSDISTLPTVPSSPDLEVSETAMEVDTPAEQFLQPTSTSTMSAQAHSTSSPTESPHSTPL 436
Db 668 SSSNSSSSVASSSSVDSSSSS-----AGVNSSSSSSAEAGSSASSSSSESSVAS 717
QY 437 LSSPDSEQRQSVASGHHHTHQSDNNNEKLSPKPGTCEPVLSLHYSTEGTITSTIKLNF 496
Db 718 SSSVDSQSSAGVNG-----SSSSSESS-----ASSNSSEGSVASSSSVDSS 762
QY 497 DEWSS--IASSRGIGSHCKSEGOEESFVPOSSVQPPPEGDSKTAPEESSEDVTKYQEGV 554
Db 763 QSSSAGVNSSSSSAEAGSSASSSSSESSIASSSSVSGSSQSSSTGVSSSSS----- 812
QY 555 SAENPVENHINITQSDKFTAKPLDSNSGERNDLNLDRSCGVPEESASSEKAKEPETSDQT 614


```

QY      756 LLEADNHVNCLOHPFPDPLASSGIDYDIKWSPEESRIFNRKLAD---EVIT--RN 809
       : |
       : |
Db      620 EKLDSTKELVPLSDTDTINSSLGNEDSIYSLDDADAISENLTDVPLMEIKTPKY 679
       : |
       : |
QY      810 ELMLEET-----RNTITVP-----ASFMLRLMASLN 835
       : |
       : |
Db      680 EVISSEVSISTVEDNVTAMPQVEYTSFPMNDPNSLN 719
       : |
       : |

```

RESULT 25

B42680

nucleolus-cytoplasm shuttle phosphoprotein - rat

N;Alternate names: Nopp140 protein B

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000

C;Accession: B42680; A42680; S27899; S30511

R;Meier, U.T.; Blobel, G.

Cell 70, 127-138, 1992

A;Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.

A;Reference number: A42680; PMID:1623516

A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm
A:Reference number: S27889
A:Accession: S27890
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:g205751; PID:AAA41719.1; PID:g205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, '0', 181-734 <ME12>
A:Cross-references: EMBL:M94287; NID:g205749; PID:AAA41718.1; PID:g205750
A:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

[illegible]

```

Db 304 SSSSESSSEEEBQKPMKKKAGPYSSVPPSVSLSKKSVGAQSPKKAQAOTQPADSSA 363
QY 504 SSSRGIGSHCKSEQEESFYQSSV--QPPEGDSETHAPEESSBDV--TKYQEGVSAENP 559
Db 364 DSS--EESDSSSEEEKTPAKTVVSKTTPAKPAPYKKAESSDSDSDSSDEFAKPK 419
QY 560 VENHINTQSDKTKAP-----LDSNGERNDLNLDRSCGVPEESASSEK----- 604
Db 420 VSAKPSLSPKPAVTPKPPPAKAVATPKQAGSGCKPQSRKADSSSEESSSEEEATKK 479
QY 605 -----AKEP-----ETSDQTSTESATNNTNTPPQFTQATGPSAH 641
Db 480 SVTTPKARVTAAPSLPAKQAPRAGDSSDSSSESSSEEEKTPPAPPAKKAAGAAMP 539
QY 642 EETSTRDSALQ-----DSTDSDDDPVLIPGARYAGPGRRSAGAVARIQEFFRRRKRKE 695
Db 540 KTPVPYKAAEASSSSSSSEDSSEEEKPKSKATPKPQAGKANGVPASONGRAGNESEE 599
QY 696 MBELDT 701
Db 600 EEE-DT 604

RESULT 26
T21315
hypothetical protein F23H12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999.
C:Accession: T21315
R:Kershaw, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19405
A:Accession: T21315
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-802 <WILL>
A:Cross-references: EMBL:Z74472; PIDN:CAA98943.1; GSPDB:GN00023; CESP:F23H12.5
A:Experimental source: clone F23H12
C:Genetics:
A:Gene: CESP:F23H12.5
A:Map position: 5
A:Introns: 24/1; 64/3; 109/1; 133/1; 522/3; 548/3; 663/3; 704/1; 740/1

Query Match 3.98; Score 176.5; DB 2; Length 802;
Best Local Similarity 21.98; Pred. No. 0.05;
Matches 101; Conservative 69; Mismatches 194; Indels 97; Gaps 21;

QY 250 NKKSCRVTSLCYSEDGQELVYSDDYILFDPKDDTARELKTPSAERREELRQ----- 304
Db 42 NLMIRVRA-CSTVRGYKCL-GHNKEFQSCNSPPKRNSLDYEDETADREMAKKOLYQDY 99
QY 305 -PPV-----KRLRLRGWS-----DTGPRAPSESERERDGEQSPNVSLMQR----- 344
Db 100 EPEIPDEAKYAQTGNRNFMHISOTTAMPAPRA-PTSSPTVSASTSDHFSVMNRGGTV 158
QY 345 --MSDMLSRNFEAEVAQSNRGRGRPRGCTSQSDISTLTPTVSSPDLE--VSETAME 400
Db 159 GMMTDEEEVFDETVKM-----SPQSPKPYTPSPVSTFEFFKPKSTEEP 203
QY 401 VDTPAEQFLQPSSTMSAQAHSTSSSTESPHSTPLISS-----PDSEQRQSEASG 452
Db 204 TSTPS-----TSTSTTTTTTSTTTTTTTEAPTTTTTKTTTSTESTEPQTVTNPSKIVD 258
QY 453 HHTHHQDNNN-----EKLSPK-----PGTGEPLVLSHYEGTITTIKLNPF- 495
Db 259 HEAFVEPDVNRCKYTLPEPLFVSVPTPAAPGTQDPDEL--EPSETPTSPSFP 316
QY 496 --TDWESSIASSKRGICSHCKSEQEESFYQSSVQPEGDSETKAPES-----SEDV 547
Db 317 IDMDPEDSTDGSHSVGY--SIGKE-----PMNGEPEVPLASNTQAPSRNWPFAISPST 370
QY 548 TKYQEGVSAENPVENHINTQSDKFTAKPLDSSNGERNDLNLDRSCGVPEESASSEKAKE 607

```

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32271
R:Murray, J.; Wohldmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid ZC178.
A:Reference number: Z21143
A:Accession: T32271
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1459 <MUR>
A:Cross-references: EMBL:AF024496; PIDN:AA870340.1; GSPDB:GN00023; CBSP:ZC178.2
A:Experimental source: strain Bristol N2; clone ZC178
C:Genetics:
A:Gene: CBSP:ZC178.2
A:Map position: 5
A:Introns: 1099/2

Query Match 3.9%; Score 176; DB 2; Length 1459;
Best Local Similarity 21.4%; Pred. No. 0.12;
Matches 143; Conservative 69; Mismatches 281; Indels 176; Gaps 31;
QY 87 TTRSGHRANIFSAKFLPCTNDQIVSCGCGVIFYT-----NVEQDAETNRQ 134
DB 274 TTAGPSSATNASSSETPCNSETQ-----TSDGTSTMTVPNDSTTAGPSSATNASSSETP 329
QY 135 QFTCHGYTYEIMTVPNDPYFLSCGEDGVRFEDTRIKTCTKEDCKDDILINCRRAA 194
DB 330 CNSETQSDGTSTMTVPNDSTT---AGPSSATN---NSASSETPCNSETQTSQSD-----G 377
QY 195 TSAVACPPPIPYLLAVGCSDSVRIYDRMLGTRATGNVAGTGTGMVARFIPSHLNKSC 254
DB 378 TSTMTVP-----NDST-----TAG-----PSSATNNSA 400
QY 255 RVTSICLYSE---DQQLIVSYSDYIYLFDPKDDT-----ARELTPSAERRE 300
DB 401 SSETPCNSETQSDGTSTMTV-----PNDSTAGPSSVTNASSSETPCNSETQ--- 449
QY 301 ELRQPPVKRLRLRGDWSDTGP-----RARPESERERDEQSPNVSLMQRMSDMLSRWF 354
DB 450 --TSDGTSTMTVSSDSTTPGSSATNASSSETPCNSETQTSQSDGTFTMTVSSD--STTAG 505
QY 355 EASEVAQSNRG---RGRSRPRGTSQSDISTLPTVPSPDLEVSETAMEYDTPAEQFLQ 410
DB 506 PSTVTNASSSETPCNSETQSDGTSTMTVSSDSTTPGSSATNASSSETPCNSETQ 563
QY 411 PST--SSTMSAQAHSTSS--PTESPSTPLLSPPDSEQSQVSEASGHHTHQSDNNNEKLS 468
DB 564 TSDGTSTMTVPNDSTTAGPSSATNASSSETPCNSETQTSQSDGTSTMT---VPNDSTTAGP 620
QY 469 -----KPGTGEPLVLSHYSTEGTTSTIKLNFDEW--SSASSSRGIGSHCKSGQEE- 520
DB 621 SSTVTNASSSETPCNSETQTSQSDGTSTMTVSSDSTTAGPSSATNASSSETPCNSETQSD 680
QY 521 --SFVPSQSVQPPGDSKTKAPESESE---DVTKYQGVSAENPVENHINITSQDKFETA 574
DB 681 GTSTMTVSSDSTTAGPSSATNASSSETPCNSETQTSQSDGTSTMT-----VSSDSTTA 732
QY 575 KPLDSNGSRNDLNLDRSCG-----VPPE-----SASSEKA--KE 607
DB 733 GP---SSTATNASSSETPCNSETQTSQSDGTFTMTVPNDSTTAGPSSVTNASSSETPCN 789
QY 608 PITSQDSTESATNENNTNPEQFQTEATGPS--AHEETSTRDSALQDPTDSDDD--PVLI 664
DB 790 TQTSQDGTSTMTVSSDSTT-----AGPSSATNASSSETPCNSETQTSQSDGTFTMTV 839
QY 665 PGARYRAGP 673
DB 840 PNDSTTAGP 848

RESULT 29
C41859

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ
C:Species: Haemophilus influenzae
A:Variety: strain HK613
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: C41859
R:Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae
A:Reference number: A41859; MUID:92234949; PMID:1373717
A:Accession: C41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1849 <POU>
A:Experimental source: strain HK613
A:Note: sequence extracted from NCBI backbone (NCBIP:97285)
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 3.9%; Score 175.5; DB 2; Length 1849;
Best Local Similarity 21.8%; Pred. No. 0.17;
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;
QY 250 NNKSCRY-----TSLCYSEDGQEILVSYSDYIYLFDPKDDTARELKTSAEERR 299
DB 921 NNKSNKVVVNVKSATGNFTLQVADKTGP-----NHNELTLFDASNATNNLEVTLANGSV 975
QY 300 EELRQPPVKRLRLRGDWS-----DTGPR---ARPESERERDEQSPNVSLMQRMSDMLSRW 352
DB 976 D-----RGAWKYKLRNVNGRYDLYNVEKRNQTVDTNITTPNDI----- 1016
QY 353 FEEASEVAQSNRGSRPRGTSQSDISTL--PTVPSPDLEVSETAMEYDTPAEQFLQ 410
DB 1017 -QADAPSAQSN-----NEELARVETPVPPAPATASAIASEQEPETRAPETAQ 1062
QY 411 P-----STSTMSAQAHSTSPSTESPSTPLLSPPDSEQSQVSEASGHHTHQSDNNNE 464
DB 1063 PAMEETNANSTETAPKSDTATQENPNS---SVPSSTETTKVAENPQNETVAKNEQE 1119
QY 465 KLSKPGTGE-----PVLSLHYSTEGTTSTIKLNFDEWSSIASSSRGIGSHCKSGEQ 518
DB 1120 ATEPTPQNGEVAKEQDPVEANTQTNATQSEK---TEETQTAETKSEPTESVTVSENQ 1176
QY 519 EESFVPSQSVQPPGDSKTKAPESESDVTKYQGVSAENPVENHINITSQDKFTAK-PL 577
DB 1177 PEKTVSQSTEDKVVVEKEEKARVE--TEETQAPQVTSKEPP-----KQAEPAPEEVP 1228
QY 578 DSNSEGRNDLNLDRSCGYPEESASSEKAKEPETSQDTSATNENNTNPEQFQTEATG 637
DB 1229 DTNAEEAQAALQQTQTTVAEAETTPSNKPAETQOPS-----EKTNAEP-----VT 1275
QY 638 PSAHEETSTRDSALQDPTDSDDDPVL-IPGARYRAGPGRRSARVARIOEFFRRRKERKEM 696
DB 1276 PVVSENTATQPTETEETAKVEKTEQVVPQVASEQSPKQEQPA-AKPOAQTKPOAPEARE 1334
QY 697 EELDTLNTRRP 707
DB 1335 NVLTTKNVGP 1345

RESULT 30
T39903
serine-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: T39903
R:Layne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z21889
A:Accession: T39903
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-534 <LYN>
A:Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13

QY 68 SGSDTKLVISNPYRKVLTIIRSHRANIFSAKFLP--CTNDKQIVSCSGDGVIFYINV 125
DB 842 SGEDTT-----VVAVESGEQPASSSTSIPTFLSKDDQVTEASGETTAAAT 890
QY 126 EQDAETNRQCOFTCHYGTITIMTPNDPYTL-----SCGEDTVKWFTRIKTSCKED 181
DB 891 EASEET-----TTSVATGESGETTVVAVESGEE-----PASSSTSIPTFL 933
QY 182 CDDILLINCRRAATVAICPPPIYLAVCCSSSVRIYDRMLCTPATCNYAGRTTGMV 241
DB 934 SKDDQVTEASGETTAAATEAS---SEETTSVATGESGETTSAVTEGSEET--- 987
QY 242 ARFISHLANKSCRTSLCYSDGQILVSYSSD-----YIYLFDPKDDTARELKT 292
DB 988 TSAVPEGNS---TTEAPAVTGSLEIPSSSESSSTTHDPISIPVITPKPVSSTIEN 1043
QY 293 PSAEERREELRPPVKRLRLRGDWSGTGPAPRPERDRDGEOSPVSIMQMSDMLSRW 352
DB 1044 VMSKTSSEAAEKKIIGEHQTGKDDAG-----KEEDNMFAFVTANPAGTSTTES 1094
QY 353 FEASEVAQSNRCGRSRPRGGTSQSDISTPLTPVSPDLEVSETAMEVDTPAEQFLQPS 412
DB 1095 AENVSTGDEENIKAKELGKQFADLAKL---AAKGVNLTETADAKS-----GE 1144
QY 413 TSSTMSAQAHSTSSPTSPHSTPLLSPPSEQRQSVASGHHTHSDNNNEKLSPKPGT 472
DB 1145 TAHVEDQVSTESSIGSEETTVNKETEE-----HHEA-SGEEDDAPAFVT 1192
QY 473 GEPVLSHSTEGTITVSTIKLN--FIDEWSSIA---SSSRGISHCKSEGESEFVPOSS 527
DB 1193 GAP-----TDSSTEASVSTSAITDETTSVAADSTSTASG-----EVQSSAIDSA 1240
QY 528 VOPPEGDSKETAP---EESSEDTV-----KYQEG---VSAENPVENHINI-TQ 568
DB 1241 TVASEGTSSEATSVIESSGEEVTTDENLVTSVAQLEGSGGITAESKDESDSVTTEAT 1300
QY 569 SDRFTAPLDSNGERNDLMDRSCGVPERASSEKAKETSDQ-----TSTESATNENN 624
DB 1301 SQSTTVSSESDSGGE-----STVAPNDSETSTTESSTTDEGSGVTAESKDEESS 1352
QY 625 TNPEPQFOTRATGSAHEETSTRDS--ALQDQDSDDDDPVLIP 665
DB 1353 TTEAPAVTISKTSSEDEEDSPTHFELGIDETMKNKSLVP 1395
RESULT 33
A48819
nuclear autoantigenic sperm protein - human
C;Species: Homo sapiens (man)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A48819
R;O'Rand, M.G.; Richardson, R.T.; Zimmerman, L.J.; Widgren, E.E.
Dev. Biol. 154, 37-44, 1992
A;Title: Sequence and localization of human NASP: conservation of a Xenopus histone-binding site.
A;Reference number: A48819; MUID:93050782; PMID:1426632
A;Accession: A48819
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-787 <OIR>
A;Cross-references: GB:M97856; NID:g184432; PIDN:AAA36027.1; PID:g184433
A;Experimental source: testis
A;Note: sequence extracted from NCBI backbone (NCBIN:117050, NCBIP:117051)
Query Match 3.9%; Score 174.5; DB 2; Length 787;
Best Local Similarity 19.5%; Pred. No. 0.063;
Matches 126; Conservative 96; Mismatches 228; Indels 195; Gaps 29;
QY 283 KDDTARELKTPSAERREELRPPVKRLRLRGDWSGTGPAPRPERDRDGEOSPVSIMQMSDMLSRW 335
DB 124 EESLVENNNDINDEARELREQVYDAMGEKAEKTEDKSLAKETDKQDSEMEKGR 183
QY 336 -----SPNVSLMQMSDMLSRWFEEASEVAQSNRCGRSRPRGGTSQSDISTPLTPVSPSP 390

DB 184 EDMDISKSAEPQEKVDLTLDMLTETSEEA-----KGGAAPE-GPNEAEVTSKPEQVEYP 237
QY 391 DLE-----VSETAMEVDTPAEQFLQPSTSTSTMSAQAHSTSSPT 428
DB 238 DAEERKSVSGTDVQRECKRGQGEQKGVIVSLEEKPEVSEEQPVVITLQKQCTAVEVEA 297
QY 429 ES--PHSTPLLSPPSEQRQSVASGHHTHSDNNNEKLSPKPGTGPVLSLHYSTEGT 486
DB 298 ESLDPTVKPDVGGDPEEKVV-----TSENBAKAVLEQLGVQVPPPAESPEVQ 348
QY 487 TWTSTIKLNTDWSSTASSRGISGSHCKSEGESEFVPO--SSVQPE--GDSETKAPEES 543
DB 349 TEAA-----EASAVEAGS---EVSERPGQAPVLPKDGAVNGSVVGDQPTPIEPQTS 397
QY 544 SEDVTYKQGVSAENPVENHINITSQDKFTAKPLDLSNGERNDLNDLDRSCGVPEESASSE 603
DB 398 IERLTETKDGSGLEEKV-----RAKLVPQSEETKL 427
QY 604 KAKEPETS-DOTSTESATNENNTNPEQTEATGSAHEETSTRDSAL---QDQDSDSD 659
DB 428 SVESSEAAAGDVTQVQAQATEKSPEDKVOI-----AANEETQEREQOMKEGETEGSEE 482
QY 660 DPVLIPGARYRAGPDRRSASAVARIQEFFRRRKRKEMEELDTLNI---RRPLVMYVKGH 716
DB 483 DD-----KENDKTEEMPNDVLENKSLQENEEEGEIGNLEAWMDLDAKIIFK-- 530
QY 717 RNSRTMIKEANFGANFVMSGDCGHFIWDRHTAHLML---LEADNHV-----VNC 766
DB 531 ---RQETKQAQAAQ-----AHLKIGEVSESVENVTQVAAEEQSC 568
QY 767 -----LOPHPPDPILASS---GIDYDIKIWSPLESRIFNRKLDADEVITRNLML 815
DB 569 LNLOEQVLEAH--DRLLAETHYQLGLAYG-----YNSQY-DEAAVQFSKSIEV 613
QY 816 TRNTITVPASFMRLMLASLHNRADLRGRSGSGGOENEDDEE 860
DB 614 IENRMV-----LNE-----QVKEAEGS-SEYKKEIEE 640
RESULT 34
G86467
hypothetical protein F7P12.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: G86467
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86467
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1587 <STO>
A;Cross-references: GB:AE005172; NID:g10092383; PIDN:AAG12790.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
Query Match 3.9%; Score 174.5; DB 2; Length 1587;
Best Local Similarity 20.0%; Pred. No. 0.16;
Matches 191; Conservative 125; Mismatches 393; Indels 245; Gaps 47;
QY 9 HLLWDVVRKSLGLEDDPSRLRSRYLGRREFIQRLKLEATLVNHDGCVNTICWNDTGEY--- 65
DB 463 HL--DVTSEGTSSVPSDEDTYITTTMEISVPLASPNVLTARDSIKTKLOSTQDFQFP 520
QY 66 ILGSGDDTKLVISNPYRKVLTIIRSG---HRANIFSAKFLPCTNDKQIVSCSGDGVIFY 122

Db 521 TMDLVEAARAEKSSCVLGKKGSGVEELGCELF-FVAAEADK--CDVRGDSLHL 577
QY 123 TNVQDAEARNOCQFCHYGTTEIMVNDPVTSLSCGEGIV-----RWFD 171
Db 578 NEISERMESS-----MSNKEDDP--FLAAKETSLPLSTDFINPETLWLYE 621
QY 172 RIKTCTKEDKDDILINCRRATSAVAICPPPIYILAVGCSDDSVRIYDRMLGTRATGN 231
Db 622 DVQAS--PEFCYSSKANAENPSS-----GCSPSTEGI-----DC 655
QY 232 YAGRTGTGMVARFIPFIPSHLNKSCRVTSLCYSDGOELLYSYSDYI-----YLDP--K 283
Db 656 FETSGCSAF-----DLAAEVESLHOE-----VSEETEFVKVMGVSPEPLG 700
QY 284 DTARELKYPSAEERE--ELQPPVKRLRLR-----GDWS-----DTGPRARPESERE 330
Db 701 ADIRSHENGESDQVIEVSAEPVAKADIQSHENGEGTESGQVIEVSPKSFSEAPT 760
QY 331 RD-----GEOSPNNVLMQSDML--SRWFEAESEVAQSNRGRSRPRGTT--S 376
Db 761 LEILTEAGAGIIGSPFELAVETESNLLHQSIGETKNEIRSHEDYGETEDYGETCS 820
QY 377 QSDISTLPT--VPSSPDLEV-----SETAME-----VDTPAEQF 408
Db 821 WPDIAVSPSSVSPPEPTLEILTDEARGLLGSEFLSEVTEIEIENLLHQSNNVETKADIL 880
QY 409 LQPSVSTMSAQAHSTSPSPSPHSTPLSSPDSEQVSAEASGHHHTHQSD-----NNNE 464
Db 881 IHEDYGETEVSQITVSPNSFSAEPTLETEDSRQ--QARGLVGSDSEFQSEVAMKTECE 939
QY 465 KLSKPGTGEVPLSLHYSTEGTSTTKLNTPTDEWSSSTASSRIGSGHCKSEGQEEFVP 524
Db 940 NLLQKRNGETKVS--SRQASPVSDCLSTPKDRLSSI--NTDDIQSLC-----983
QY 525 QSSVQPPGEGSETKAPESSEDVTKYQEGVSAEN--PVENHINITOSDKFTAKPLDSNS 581
Db 984 -SSQPP--SESEVPATQD--QESGIITEKPKTELLIGSGSEKYS--LSEIE 1034
QY 582 GERN-DLNLDRSCGVPEESASSEKAKPEPTSDQSTSEATNNTNPEPQFTEATGPSA 640
Db 1035 GEENTDGLRSLRCPISALAAKTSQEDPKLIEELSSDSSGSGQENQ-----PET 1082
QY 641 HETSTRDQLQTDSDDDPVLIPGARYRAGPDRRSARVARIQEFRRKEREKEMEELD 700
Db 1083 H--AVRDDVLCDMSSTCNI-----WSRGRKAASVLKRTNKSQKQKQ--1125
QY 701 TLNIRPLKVMYKGRNSRTMIKEANF--WGANFVMSGDCGHIFIDRHTAEHLMLL 757
Db 1126 --TGROPKDKL--HRKQALSKSISLTHHGAELP-----EFTPDKEN--L 1168
QY 758 EADNHVNCIQLPHFPDIPILASSGIDYDIKIWSPLESRIFNRKLADDEVITRNLMLLEPR 817
Db 1169 TPSSHMLKRLQDIG-DVKDSKSLKSGKSCSLVHSSI--AVLASEATEPEIFTDK 1225
QY 818 NITVPASFLMLASLNHADRLEGDRSGEGQ-----ENENED 858
Db 1226 N--LTPSSHMLKRLRFBGDIKDTK--GSSSKATRKPFEDIRMEENVYVQEPED 1275

RESULT 35

A46194
neurofilament protein NF-220, high-molecular-weight splice form - longfin squid
C:Species: Loligo pealeii (longfin squid)
C:Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Mar-2000
C:Accession: A46194
R:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Gainer, H.; Battey, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992
A:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod
A:Reference number: A46194; MUID:92357751; PMID:1379729
A:Accession: A46194
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA

A:Residues: 1-1200 <WAY>
A:Cross-references: GB:M94389; NID:g161291; PID:g161292
A:Experimental source: stellate ganglion
A>Note: sequence extracted from NCBI backbone (NCBIP:113499)
C:Superfamily: neurofilament triplet H protein
C:Keywords: alternative splicing

Query Match 3.8%; Score 174; DB 2; Length 1200;
Best Local Similarity 19.7%; Pred. No. 0.12;
Matches 105; Conservative 79; Mismatches 220; Indels 130; Gaps 18;
QY 282 PKDDTARELKTSAERREELRQPPVKRLRLRGDMSDTGPRAR-----324
Db 682 PLSDT--IKSPVSE-----PAISPVSSVSKGASPTSPARTMTPIGSGSEKSAKSPV 731
QY 325 -----PESERERDEQSPNVSLMQMSDMLSRWFEAESEVAQSNRGRSR--RPRGTT 376
Db 732 RSEATKSPVSEKSGSKSPVSEALSPVLSSESVHSTAMSHTSRSTASEKSVKSP 791
QY 377 QSDISTLPTVPSPDLEVSETAMEVDTPAEQFLOPSTS--STMSAQAHSTSS-----PTE 429
Db 792 HSERTASPTAKSP-----IIMEPAKSPKDESEKELSPERSEVMGSKSQITSSSAKSPVSE 848
QY 430 SPHSTPLSSPDSEQVSAEASGHHHTHQSDNNNEKLSKPKGTPGVLS--LHYSTEGTTT 488
Db 849 KADSEKATSPTSPSEKVDDESSARSPTHSRSEKOKSARSPMTSDHIKSPIDFEKAESK 908
QY 489 STIKLNFTE-----WSSIASSSRGIGSHCKSEGOEE-----SFVPQSSVQPPGDS 535
Db 909 SALSYKSDHDKSPVSEKAESEKARSPPVSEKAESEKARSPTSEPAKSPVSEKAA 968
QY 536 ETKAPEESSEDVTKYQEGVSAENPVENHINITOSDKFTAKPLDSNGE-----RNDLMLD 590
Db 969 ESEKARSPPVSEKAESEKARSPPVSE--EKAESKARSPPVSEKDESEKSAKSPSTDQ 1026
QY 591 RSCGVPEESASEKA-----KETSDQSTSEGA-----TNENTNPEQFQTE- 634
Db 1027 ARSPVSEKAESEKARSPTVSEHVKSPPVSEKSEKARSPTVSEHVKSPPVSEKAE 1086
QY 635 -----ATGPSAHEET-----STRDSALQD 655
Db 1087 EKSAPVSEPAKSPVSEKAESEKSAKSPVSEHATSPVASEKSEKSGKSPASEKA 1146
QY 656 DSDDPVLIP-----GARYRAGPDRRSARVARI-----QEFFRRKEREKEMEELD 700
Db 1147 ESEKSPVSEKSGSKTAGSGSDGIIITTTTSOERRSMSEKSKMSGSSD 1200

RESULT 36

T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GE>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match

Best Local Similarity 3.8%; Score 174; DB 2; Length 232;
Matches 88; Conservative 37; Mismatches 188; Indels 50; Gaps 9;

Db 542 RPRTRSAKAELEENQSRKQKRRRIKQVEDSSSENKSNSEEEEEK -----EBEEE 595
QY 646 TRDSALQDTDDDDPVLIPGARYRAGPGRDRSAVARI -----Q 684
Db 596 EEEEEEEDENDDD -----SKSPGKGRKKIRKILKDDKRLTETQNALKEEERRK 646
QY 685 EFFRRKRRKMEEL -----DTLIRPLVVMVYKGRHNSRTWIK 724
Db 647 RIAEREREKRELVIEIEDASPTKPTIKLVLDDEDEETKEPLVQV -----HRMVIK 702
QY 725 EANFWGANFVMSGDC 740
Db 703 PHQVDGVQFMW ---DC 715
RESULT 39
B98002
Iga-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae.
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: B98002
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: B98002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1963 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:915458662; GSPDB:GN00174
A:Gene: iga
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
Query Match 3.8%; Score 173; DB 2; Length 1963;
Best Local Similarity 18.1%; Pred. No. 0.25;
Matches 146; Conservative 120; Mismatches 299; Indels 242; Gaps 29;
QY 111 IVSCSGDGVFTYNNVQDAETNRQCQFTCHYGTYYEY -----MTVPND--PYTFLSCGED 163
Db 139 LVAGMNGVLVSVHALEN -----HLLNNTDYELTSGEKLPLPKREISYTYIGYKE 191
QY 164 G-----TWTFTRIKTSCTK ----- 179
Db 192 GKTTSDFEVSNOBKSAATPTKQKVDYNTVNFVDHPSTVQAIQEQTPVSSSTKPTQVQV 251
QY 180 EDCDDILINCRRAATSAICPPPIYPYLVAGCSDSSVRIYDRMLGSDTGPRAPESEERERDGG 239
Db 252 EKPFTSELINPREEKQ -----SDSQEQLAEHKNLETKKEEKISPKETG 297
QY 240 MVARFTPSHLNKSRCVTSICYSDEQOELVSYSS ----- 274
Db 298 VNTLNPQDELSQLNKKPELLYREETIETKIDQEEIQENPDLAEGTVRVYKQEGKLGKV 357
QY 275 DYIYLFD-PKDDTARELKTPSAEEREELRQPPVKRLRLRGDWSDTGPRARPESEERERDGG 333
Db 358 EIVRFYSNKEVSRIVSTST-----TAPSPRIVEKGTGKTQVIRKQPETGVHEKD 409
QY 334 EQS-----PNVSLMRMSDLSRWFEASEVAQSNRGRSRPRGTSQSDISTLPTV-- 386
Db 410 VQSGAIVEPAIQ-----PELPEAVVSDKGEVEVQ-----TLPEAVV 446
QY 387 -----PSPDLEVSETA-----MEVDTAPQFLQPTSTSTMSAQ 420
Db 447 TDKGTEVQEPSPDVTWSDKGEPEQVAPLPEYKGNIEQVKPETPVETKEQGEKTEEV- 505
QY 421 ABSTSSPTESPHSTPLLSPPDEQSQSVASGHHHT--HQSDNNNEKLSPKPGTGEPLVS 478
Db 506 -----PVKPEETVPNNEGTEGTSIQEAENPVQPAEESTTNSEKVP----- 549

QY 479 LHYTEGTTTSTIKLNTFDWSSIASRRGIGHCKSEQEBSFVQSSVQPPGDSQETK 538
Db 550 ---DTSSNTGEVSSNPSTSTSVGSENKPEHNSKNSEKT-VVEVPVNEGIVEGT 605
QY 539 APESSSDVTYQGVSAENPVENHINTOSDKFTAKPLDLSNGSRNDLNLDRSCGVPEE 598
Db 606 SNQETEPVQPAEE-----TQTN--SGKTANENTGEVSNKPSDKPPV-EE 648
QY 599 SASSEK---AKEPETSQDTSTESATNE-----NNTNPEPQFOTATGSPSAHETSTR 647
Db 649 SNOPEKNGTATKPNSENGTTSNGQTEPEKKLELNVSDIELYSQTNGT-----YRQHVSL 704
QY 648 DSALQDTD-----DSDDDPVLIPGARYRAGPGRDRSA-----VARIQEFFRRR--- 690
Db 705 DGIPENTDTTFVKVKVSAFKVDYIPVASITEEKRNGQSVYKITAKAEKLQLEKNKYVDN 764
QY 691 ----KERKEEELDTLNRRLPLVMVYKGRHNSRTMIKEANFWGANFVMSGDCGHIFTW 746
Db 765 FTFLDKKAKEENTFTSFSNLVKAINQNPSTGYHLAASLN---ANEVELGPD-ERSYIK 820
QY 747 DRHTAEHLMLLEADNHNVCLOPHFED 773
Db 821 DTFTGLRIGEDKGNKYAIYNLKKPLFE 847
RESULT 40
T30989
serine/threonine protein kinase NIK - mouse
N:Alternate names: Nck interacting kinase
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T30989
R:Su, Y.-C.; Han, J.; Xu, S.; Cobb, M.; Skolnik, E.Y.
EMBO J. 16, 1279-1290, 1997
A:Title: NIK is a new Ste20-related kinase that binds NCK and MEK1 and activates the
A:Reference number: Z20954; MUID:97280817; PMID:9135144
A:Accession: T30989
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1233 <SU>
A:Cross-references: EMBL:U88984; NID:g1872545; PID:g1872546; PIDN:AAC53165.1
C:Keywords: protein kinase
Query Match 3.8%; Score 172.5; DB 2; Length 1233;
Best Local Similarity 20.8%; Pred. No. 0.14;
Matches 139; Conservative 75; Mismatches 261; Indels 193; Gaps 29;
QY 283 KDDTARELKTPSAEERRELRQPPVKRLRLRGDWSDTGPRARPESEERERDGEQSPNVSLM 342
Db 409 EQQRREAREARQREQRREQEKRL-----EELERKEEERERRAEERKRV--- 460
QY 343 ORMSDMLSRWFEAS---EVAQS-----NRGRSRPRGGTQSQSDISTLPTV 386
Db 461 EREQYIIRQLREEQRHLETLQQLLQEQAMLHDDRPHAOQQPPPPQQQDRSKPSFHA 520
QY 387 PS-SPDLEVSETAMEV-----DTPAEQFLQ-----PSTS 414
Db 521 PEPKPHYDPADRARREVQWHLASLANNVSPVSRSHSFSDPSKFAHHHLRSQDPCPPSR 580
QY 415 STMSAQAHSTS---SPT-----ESPSTPLLS---SPDSEQRQSVASGHHHTHQ 458
Db 581 EGLSQSDSKSEVPEPTQKAWRSDSDEVPVPVPTTSRSPVLSRRDPLQGGGQNSQ 640
QY 459 SDNNN-----EKLSPKPGTGEPLVSLHYSTEGTTTSTIKLNTFDWSSIAS 504
Db 641 AQQRNSTRSSIEPRLWVERKLVPRPGS-----SSSGS 674
QY 505 SSRGI---GSHCKSE-GQESFVQSSVQ-----PPEGDSQETKAPESSE--- 545
Db 675 SNSGSGQPSGHPGSGGGERFRVRSRSGSESPSPQSAAKKPDCKKEVFRSLKPAGEV 734
QY 546 DVTYQGVSAENPVENHINT--QSDKFTAKPLDLSNGSRNDLNLDRSCGVPEE--SAS 601

Db 735 DLTALAKELRAVEDVRPPHKVTDYSSSSSESGTTDEEEDVEQEGADDSTSGPEDTRAAS 794
QY 602 SEKAKPEPETSOTSTESATNENNTNPEQOTEAUGPSAHEETSTRDSALQDTDDSD-- 659
Db 795 SPNLSNGETE---SVKTMIVHDDVESEPAWTPSKEGTLIVROTSASSTLQKHKSST 851
QY 660 ----DPVLI-----PGARYRAGPCD--RRSAVARIQEFFRRKERRKEM 696
Db 852 PFIDPRLLQISPSSGTTVTSVVGFSCDGLRPEAIRQDPTKGSVVNVNPTNTR----- 904
QY 697 EELDTLNIRPLVKVYKGRHNSRTMIKEANFWGVMSGDCGHIFWDRHTAHLML 756
Db 905 PQSDTPEIRK-----YKKRFNSEILC--AALWGVNLLV-GTESG-----LML 943
QY 757 LE--ADNHVVNCLQPHFPDPIASSGIDYDIKIWSPLIESRIFNRKLADEVITRNELMLE 814
Db 944 LDRSGQGVVPLISRRRFQOMVLEGLNVLVTISGKKDKLRVYLSWLRNKLHNDPEVE 1003
QY 815 ETRNTTIV 822
Db 1004 KKGWTTV 1011

Search completed: March 18, 2003, 15:36:23
Job time : 76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 25, 2003, 20:18:38 ; Search time 7524 Seconds
(without alignments)
9979.431 Million cell updates/sec

Title: US-09-781-693A-1
Sequence: 1 atgtctcgggggtggtctccta.....atgaataatgagatgagaa 2580

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sv:*

13: gb_vt:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_cm:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hgt_hum:*

31: em_hgt_inv:*

32: em_hgt_other:*

33: em_hgt_mus:*

34: em_hgt_pln:*

35: em_hgt_rod:*

36: em_hgt_nam:*

37: em_hgt_vrt:*

38: em_sv:*

39: em_hgt_hum:*

40: em_hgt_mus:*

41: em_hgtg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2573.6	99.8	3189	6	AX076993	AX076993 Sequence
2	2573.6	99.8	3189	6	AX113966	AX113966 Sequence
3	2508.4	97.2	3022	6	AX078303	AX078303 Sequence
4	2332.6	90.4	3420	6	AX076995	AX076995 Sequence
5	2332.6	90.4	3420	6	AX113968	AX113968 Sequence
6	2140	82.9	2448	6	AX060956	AX060956 Sequence
7	2140	82.9	2448	6	AF124434	AF124434 Homo sapi
8	2076.4	80.5	2325	9	AF150734	AF150734 Homo sapi
9	2073	80.3	2727	9	AF150734	AF150734 Homo sapi
10	1594.4	61.8	2024	9	BC025262	BC025262 Homo sapi
11	1446.4	56.1	1952	9	BSM01706	BSM01706 Homo sapi
12	1378.4	53.4	2803	9	AK093970	AK093970 Homo sapi
13	1287.2	49.9	162314	9	AC016484	AC016484 Homo sapi
14	1285.6	49.8	159098	2	AC016484	AC016484 Homo sapi
15	1087.4	41.4	1726	9	AK000984	AK000984 Homo sapi
16	567.4	22.0	65517	2	AC107981	AC107981 Homo sapi
17	328.8	12.7	48438	9	HS702319	HS702319 Homo sapi
18	328.8	12.7	163230	2	AC026140	AC026140 Homo sapi
19	266.6	10.3	74729	9	HS703H14	HS703H14 Human DNA
20	242.2	9.4	251	11	G05536	G05536 human STS
21	217.4	8.4	158324	2	AC069147	AC069147 Homo sapi
22	217.4	8.4	158324	2	AC069147	AC069147 Homo sapi
23	202.8	7.9	163307	2	AC130131	AC130131 Rattus no
24	182.6	7.1	179269	2	AC116374	AC116374 Mus muscu
25	166.6	6.5	100694	2	AC127803	AC127803 Rattus no
26	166.6	6.5	168307	2	AC130131	AC130131 Rattus no
27	159.6	6.2	179269	2	AC116374	AC116374 Mus muscu
28	142.4	5.5	100694	2	AC127803	AC127803 Rattus no
29	125.2	4.9	59143	2	AC100993	AC100993 Mus muscu
30	125.2	4.9	59143	2	AC069147	AC069147 Homo sapi
31	98	3.8	158324	2	AC069147	AC069147 Homo sapi
32	96.4	3.7	102258	9	HS295C6	HS295C6 Human DNA
33	76	2.9	1765	8	AY120722	AY120722 Arabidops
34	71.2	2.8	2594	8	AY039964	AY039964 Arabidops
35	70.6	2.7	2034	6	AX370488	AX370488 Sequence
36	70.6	2.7	2966	9	AK023101	AK023101 Homo sapi
37	70.6	2.7	3267	9	AK023778	AK023778 Homo sapi
38	70.6	2.7	4211	6	AX370510	AX370510 Sequence
39	70.6	2.7	4305	9	BC032523	BC032523 Homo sapi
40	69.4	2.7	798	9	HS295C061	HS295C061 Sequence
41	66.8	2.6	84499	8	ATT12J5	ATT12J5 Homo sapi
42	66.8	2.6	195165	8	ATCHRIV82	ATCHRIV82 Arabidops
43	66.6	2.6	114897	2	AP003624	AP003624 Oryza sat
44	66.6	2.6	141534	2	AP003685	AP003685 Oryza sat
45	65	2.5	110980	8	ATF9K21	ATF9K21 Arabidops
46	64.2	2.5	2834	9	AK098114	AK098114 Homo sapi
47	63.8	2.5	1887	6	AX370484	AX370484 Sequence
48	63.8	2.5	2206	6	AX370508	AX370508 Sequence
49	63.8	2.5	2286	3	AY059458	AY059458 Drosophil
50	63.8	2.5	78054	3	AC004296	AC004296 Drosophil
51	63.8	2.5	120625	2	AC017563	AC017563 Drosophil
52	63.8	2.5	136164	3	AC099006	AC099006 Drosophil
53	63.8	2.5	162492	3	AC099039	AC099039 Drosophil
54	63.8	2.5	255386	3	AE003800	AE003800 Drosophil
55	63.4	2.5	81631	9	HS745114	HS745114 Human DNA
56	63	2.4	81740	9	AL663123	AL663123 Homo sapi
57	61	2.4	1731	6	AX370534	AX370534 Sequence
58	61	2.4	5091	9	AB028960	AB028960 Homo sapi
59	60.8	2.4	1596	8	AY084990	AY084990 Arabidops
60	60.8	2.4	91938	8	ATF20M13	ATF20M13 Arabidops
61	60.8	2.4	93760	8	ATF22M13	ATF22M13 Arabidops
62	60.8	2.4	107740	8	AC069558	AC069558 Genomic S
63	60.8	2.4	199789	8	ATCHRIV89	ATCHRIV89 Arabidops
64	60.2	2.3	773	8	AX187045	AX187045 Sequence
65	60	2.3	773	8	AY086855	AY086855 Arabidops

Db 716 GATTGTAAGAGATATTTAAATTAACGTGCGAGCTGTCGACGCTGTTGCTATTTGC 775
Qy 601 CCACCAATACCAATATACCTTGTGTTGGTTGTTCTGACAGCTCAGTACGAATATATGAT 660
Db 776 CCACCAATACCAATATACCTTGTGTTGGTTGTTCTGACAGCTCAGTACGAATATATGAT 835
Qy 661 CGCGAATGCTGGCAGACAGAGCTACAGGAATATACAGGTGAGGAGCTACTGGGAATG 720
Db 836 CGCGAATGCTGGCAGACAGAGCTACAGGAATATACAGGTGAGGAGCTACTGGGAATG 895
Qy 721 GTTGCCCGTTTATCTCCCTTCCCAATTAATAAAGTCCGAGAGTACATCTCTGTGT 780
Db 896 GTTGCCCGTTTATCTCCCTTCCCAATTAATAAAGTCCGAGAGTACATCTCTGTGT 955
Qy 781 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTACTCTTCAGATTACATATATCTTTT 840
Db 956 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTACTCTTCAGATTACATATATCTTTT 1015
Qy 841 GACCCGAAAGATGATACAGCAGAGAACTTAAACCTCTCTGCGGAAGAGAGAGAA 900
Db 1016 GACCCGAAAGATGATACAGCAGAGAACTTAAACCTCTCTGCGGAAGAGAGAGAA 1075
Qy 901 GAGTTGCGACACACAGCTTAAAGCTTTGAGACTTCGTGTGATTTGTTGATCTGGA 960
Db 1076 GAGTTGCGACACACAGCTTAAAGCTTTGAGACTTCGTGTGATTTGTTGATCTGGA 1135
Qy 961 CCCAGCAAGCCGAGAGTGAACGAGACAGAGATGAGAGCAGAGTCCCAATGTCA 1020
Db 1136 CCCAGCAAGCCGAGAGTGAACGAGACAGAGATGAGAGCAGAGTCCCAATGTCA 1195
Qy 1021 TTGATGAGAGATGCTGTGATGTTATCAAGATGTTTGAAGAGCAGAGTGAATCAATCA 1140
Db 1196 TTGATGAGAGATGCTGTGATGTTATCAAGATGTTTGAAGAGCAGAGTGAATCAATCA 1255
Qy 1081 CAAAGCAATAGAGCAGAGAGTCTCGACCCAGAGAGTGAACAGTCAATCAATCAATCA 1200
Db 1256 CAAAGCAATAGAGCAGAGAGTCTCGACCCAGAGAGTGAACAGTCAATCAATCAATCA 1315
Qy 1141 TCAACTCTTCTAGGTCCTTCAAGTCTGATTTGAAGTGAAGTGAACAGTCAATCAATCA 1375
Db 1316 TCAACTCTTCTAGGTCCTTCAAGTCTGATTTGAAGTGAAGTGAACAGTCAATCAATCA 1400
Qy 1201 GTAGATCTCCAGTGAACAAATTTCTCAGCCTTCTACATCTCTACATCTCTACATCT 1260
Db 1376 GTAGATCTCCAGTGAACAAATTTCTCAGCCTTCTACATCTCTACATCTCTACATCT 1435
Qy 1261 GCTATCTGACATCTCTCCACAGAAAGCCCTCATCTACTCTCTCTCTCTCTCTCT 1320
Db 1436 GCTATCTGACATCTCTCCACAGAAAGCCCTCATCTACTCTCTCTCTCTCTCTCT 1495
Qy 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCTCTGAT 1380
Db 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCTCTGAT 1555
Qy 1381 ACATTAATGAAAGTGTGACCCCAACAGGACAGGAGTGAACAGTCTTAAAGTTGAC 1440
Db 1556 AACAATAATGAAAGTGTGACCCCAACAGGACAGGAGTGAACAGTCTTAAAGTTGAC 1615
Qy 1441 TACAGCAGAGAGGACACTACAGCACAATAAAGTGAATTTACAGATGAATGGAGC 1500
Db 1616 TACAGCAGAGAGGACACTACAGCACAATAAAGTGAATTTACAGATGAATGGAGC 1675
Qy 1501 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTTGCAATCTGAGGTCAGGAGAA 1560
Db 1676 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTTGCAATCTGAGGTCAGGAGAA 1735
Qy 1561 TCTTTTCGTCACAGAGTCTAGTCAACACAGAGAGAGTGAACAGTCTTAAAGTTGAC 1620
Db 1736 TCTTTTCGTCACAGAGTCTAGTCAACACAGAGAGAGTGAACAGTCTTAAAGTTGAC 1795
Qy 1621 GAAGAAATCATCAGAGGATGTCAGAAAAATATCAGGAGAGATCTGCGAGAAACCCAGTT 1680
Db 1796 GAAGAAATCATCAGAGGATGTCAGAAAAATATCAGGAGAGATCTGCGAGAAACCCAGTT 1855

Qy 1681 GAGAACCATATCAATATATAACAATCAGATAGTTTACAGCCAGCCATTTGATTCCAAAC 1740
Db 1856 GAGAACCATATCAATATATAACAATCAGATAGTTTACAGCCAGCCATTTGATTCCAAAC 1915
Qy 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1800
Db 1916 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1975
Qy 1801 TCATCTCAAAAAGCCCAAGGACCAAGAACTTCAGATCAGACTAGCAGTGTGCTAC 1860
Db 1976 TCATCTCAAAAAGCCCAAGGACCAAGAACTTCAGATCAGACTAGCAGTGTGCTAC 2035
Qy 1861 AATGAAATTAACACCTCTGAGCCCTCAGTTCACAAACAGAGCCACTGGGCTTTCAGCT 1920
Db 2036 AATGAAATTAACACCTCTGAGCCCTCAGTTCACAAACAGAGCCACTGGGCTTTCAGCT 2095
Qy 1921 CATGAAGAAACATCCACAGGAGCTCTGCTCTTTCAGGACACAGATGATGATGAC 1980
Db 2096 CATGAAGAAACATCCACAGGAGCTCTGCTCTTTCAGGACACAGATGATGATGAC 2155
Qy 1981 CCAGTCTCTGATCCAGGCTGCAAGGTATTCGAGCAGGAGCTGGTGTATGATGATGAT 2040
Db 2156 CCAGTCTCTGATCCAGGCTGCAAGGTATTCGAGCAGGAGCTGGTGTATGATGATGAT 2215
Qy 2041 GCCCGTATTCAGAGTCTTTCAGAGCGAGAGAAAGAAAGAAATGGAATTTGGAT 2100
Db 2216 GCCCGTATTCAGAGTCTTTCAGAGCGAGAGAAAGAAAGAAATGGAATTTGGAT 2275
Qy 2101 ACTTTTGAACATFAGAGGCGCTAGTAAAATGTTTATAAGGCCATCGCAACTCCAGG 2160
Db 2276 ACTTTTGAACATFAGAGGCGCTAGTAAAATGTTTATAAGGCCATCGCAACTCCAGG 2335
Qy 2161 ACAATGATAAAGAGCCAAATTTCTGGGTGCTAACTTTTGAATCAGTGTCTGACTGT 2220
Db 2336 ACAATGATAAAGAGCCAAATTTCTGGGTGCTAACTTTTGAATCAGTGTCTGACTGT 2395
Qy 2221 GGCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280
Db 2396 GGCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2455
Qy 2281 AATCATGTGGTAAACTGCTGACGACCATCGCTTTGACCCCAATTTTACCTCATCTGCG 2340
Db 2456 AATCATGTGGTAAACTGCTGACGACCATCGCTTTGACCCCAATTTTACCTCATCTGCG 2515
Qy 2341 ATAGATTATGACATAAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGAAA 2400
Db 2516 ATAGATTATGACATAAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGAAA 2575
Qy 2401 CTTGCTGATGAAGTTATTAACTCGAAACGAACTCATGCTGAGAGAGTCAAGGATTTTAAACCGAAA 2540
Db 2576 CTTGCTGATGAAGTTATTAACTCGAAACGAACTCATGCTGAGAGAGTCAAGGATTTTAAACCGAAA 2635
Qy 2461 ACAGTTCCAGCCCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520
Db 2636 ACAGTTCCAGCCCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2695
Qy 2521 CGTTTGGAGGTCAGAGTACAGAGGCTCTGCTCAAGAGATCAAAATGAGGATGAGGAA 2580
Db 2696 CGTTTGGAGGTCAGAGTACAGAGGCTCTGCTCAAGAGATCAAAATGAGGATGAGGAA 2755

RESULT 2

AX113966

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AX113966

Sequence 1 from Patent EP1106690.

AX113966

AX113966.1 GI:13940140

3189 bp

DNA

linear

PAT 01-MAY-2001

Wed Mar 26 09:38:48 2003

REFERENCE 1 (bases 1 to 3189)
AUTHORS el Shami,A.S., Menon,S.N. and French,C.K.
TITLE Polynucleotide encoding autoantigens associated with endometriosis
JOURNAL Patent: EP 1106690-A 1 13-JUN-2001;
DIAGNOSTIC PRODUCTS CORPORATION (US)
FEATURES
source Location/Qualifiers
1. 3189
/organism="Homo sapiens"
/db_xref="taxon:9606"
176..2758
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37867.1"
/db_xref="GI:13940141"
translation="MSRGGSYPHLLMDVRKRFLEDPRLSRRLRYLGRFFIOLKLE
ATLNVHDCVNTICWNTDGEYILSGSDTKLVISNYSRKVLTIIIRSHRANFSAKF
LPCTNDQIVSCSDGVIFVTNVEQDAETNRQCFCHYGTYEIMVNPDPFFLSC
GEDCTVRFWETRIKTSCTKEDCKDDILINCRRAATVAICPPPIYLAVGSDSSVRI
YDRMLCTVRATGVAGTGMVARFPTSHLNKNSCRVTSICYSDEGQELVYSDDY
IYLFDPKDDTARELKTPSAREERELRQPPVKRLRLGWSDTGPRAPSERERDGE
QSPNYSMLQMSMLSRWRFEASEVAQSNRGRSRPRGCTSDSDISTLPVSPDOL
EVSTAMEVDTPAEQFLOPSTSTWSAQAHSTSPTESPHSTPLLSPEQSQSVEA
SGHHTHQSDNNNEKLPKPGTGEPLVSLHSTEGTITSTIKLNFIDWESSIASSESG
IGSHCKSQGESVPOSSVQPPREGDSETPAESSEDVTKYQBGVSAENPVENHINI
TQSKFTAKPLDSNGRNDLNRSCGVPEESASEKAKEPETSDQSTSTESATNENN
TNPEQFOTEGTSAHRETSTRDSALQDVTDDSDPVLIPGARYRAGPDGRKSAVAR
IQEFPRFKERKEMEELDTLNRPLVMYVKGHRNSRTMIKEANFGANFVMTGSEC
GHIFTWDRHTAHEMLLEADNHNVCLOPHFDPILASSGIDYDIKWSPLEESRIFN
RKLDADEVITRNMELMEETRNITVPASEFMLRLASLNHRIADRLEGRSSESGGENE
EDEE"

BASE COUNT 977 a 661 c 724 g 827 t

ORIGIN

Query Match 99.8%; Score 2573.6; DB 6; Length 3189;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2576; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGCTCGGGGTGGCTCCTACCCACACCTGTTGTGGGAGCTGAGGAAAGTCCCTCGGG 60
DB 176 ATGCTCGGGGTGGCTCCTACCCACACCTGTTGTGGGAGCTGAGGAAAGTCCCTCGGG 235
QY 61 CTGGAGGACCCGTCGGGCTGGGAGTCGCTACCTCGGGAAGAGAAATTTATCCAAAGA 120
DB 236 CTGGAGGACCCGTCGGGCTGGGAGTCGCTACCTCGGGAAGAGAAATTTATCCAAAGA 295
QY 121 TTAAGCTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTTAATACATCTCTTGGAT 180
DB 296 TTAAGCTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTTAATACATCTCTTGGAT 355
QY 181 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTAAGTAATAGTAACTCT 240
DB 356 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTAAGTAATAGTAACTCT 415
QY 241 TACAGCAAGAAAGTTTGAACAATTCGTTTCAGGGCACCGAGCAACACATATTAGTGCA 300
DB 416 TACAGCAAGAAAGTTTGAACAATTCGTTTCAGGGCACCGAGCAACACATATTAGTGCA 475
QY 301 AAGTTCTTACCTGTACAAATGATAACAGATTGTATCTCTGCTCGAGATGGAGTAATA 360
DB 476 AAGTTCTTACCTGTACAAATGATAACAGATTGTATCTCTGCTCGAGATGGAGTAATA 535
QY 361 TTTTATACCAAGTTGAGCAAGATGCAGAAACCAACACAGACAATGCCAATTTACGTGCAT 420
DB 536 TTTTATACCAAGTTGAGCAAGATGCAGAAACCAACACAGACAATGCCAATTTACGTGCAT 595
QY 421 TATGGAATCTATTGAGATTATGACTGTACCGAATGACCCCTTACACTTTTCTCTCTGT 480
DB 596 TATGGAATCTATTGAGATTATGACTGTACCGAATGACCCCTTACACTTTTCTCTCTGT 655
QY 481 GGTGAAGATGGAACCTGTTAGTGGTTTTCATACAGCATCAAAACCTAGCTGCACAAAAGAA 540
DB 656 GGTGAAGATGGAACCTGTTAGTGGTTTTCATACAGCATCAAAACCTAGCTGCACAAAAGAA 715

QY 541 GATTGTAAAGATGATATTTTAACTCTCGACGTGTCGCCACGTCGTGCTGCTATTTCG 600
DB 716 GATTGTAAAGATGATATTTTAACTCTCGACGTGTCGCCACGTCGTGCTGCTATTTCG 775
QY 601 CCACCAATACCATTATACCTTCTGCTGTTGTTCTTGCAGCTCAGTACGAAATATATGAT 660
DB 776 CCACCAATACCATTATACCTTCTGCTGTTGTTCTTGCAGCTCAGTACGAAATATATGAT 835
QY 661 CGCGGAATCTGGGCAACAAGAGCTACAGGGAATATATGAGTTCGAGGGAGTACTCTGGAATG 720
DB 836 CGCGGAATCTGGGCAACAAGAGCTACAGGGAATATATGAGTTCGAGGGAGTACTCTGGAATG 895
QY 721 GTTGCCTGCTTTATTCCTTCCCATCTTATATATATAGTCTGTCAGAGTGCATCTCTGTGT 780
DB 896 GTTGCCTGCTTTATTCCTTCCCATCTTATATATAGTCTGTCAGAGTGCATCTCTGTGT 955
QY 781 TACAGTGAAGATGCTCAAGAGATCTCTGTTAGTACTTCTCAGATTACATATATCTTTT 840
DB 956 TACAGTGAAGTGTCAAGAGATCTCTGTTAGTACTTCTCAGATTACATATATCTTTT 1015
QY 841 GACCCGAAAGATGATACAGCAGCAGAACTTAAACCTCTTCTGCGGAAGAGAGAGAA 900
DB 1016 GACCCGAAAGATGATACAGCAGCAGAACTTAAACCTCTTCTGCGGAAGAGAGAGAA 1075
QY 901 GAGTTGCGCAACACCACAGTTAAAGCGTTTGAAGACTTCTGCTGGTGAATTTGGTCAGATACTGA 960
DB 1076 GAGTTGCGCAACACCACAGTTAAAGCGTTTGAAGACTTCTGCTGGTGAATTTGGTCAGATACTGA 1135
QY 961 CCCAGAGCAAGCGGGAGAGTGAACGAGACAGATGAGAGCAGAGTCCCAATGTGCA 1020
DB 1136 CCCAGAGCAAGCGGGAGAGTGAACGAGACAGATGAGAGCAGAGTCCCAATGTGCA 1195
QY 1021 TTGATGACAGAAATGCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAAGTTGCA 1080
DB 1196 TTGATGACAGAAATGCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAAGTTGCA 1255
QY 1081 CAAAGCAATAGAGCAGGAGGAAGATCTCGACCCAGAGGTGGAACAGTCAATCAGATATT 1140
DB 1256 CAAAGCAATAGAGCAGGAGGAAGATCTCGACCCAGAGGTGGAACAGTCAATCAGATATT 1315
QY 1141 TCACCTCTTCTACGGTCCCATCAAGTCTGATTTGGAAGTGAAGTGAACCTGCAATGAA 1200
DB 1316 TCACCTCTTCTACGGTCCCATCAAGTCTGATTTGGAAGTGAAGTGAACCTGCAATGAA 1375
QY 1201 GTAGATCTCCAGCTGGAACAATTTCTCAGCCTTCTACATCTCTACATCTCAGCTCAG 1260
DB 1376 GTAGATCTCCAGCTGGAACAATTTCTCAGCCTTCTACATCTCTACATCTCAGCTCAG 1435
QY 1261 GCTCATTCGACATATCTCCACAGAAAGCCCTCATTTCTACTCTTCTGCTATCTTCTCCA 1320
DB 1436 GCTCATTCGACATATCTCCACAGAAAGCCCTCATTTCTACTCTTCTGCTATCTTCTCCA 1495
QY 1321 GACAGTGAACAAGGAGCTGCTGTTGAGGCATCTGACACACACACATCATCAGTCTGAT 1555
DB 1496 GATAGTGAACAAGGAGCTGCTGTTGAGGCATCTGACACACACACATCATCAGTCTGAT 1555
QY 1381 AACATAATGAAGAAGCTGAGCCCAACAGGAGGAGTGAACCAAGTTTAAAGTTTGAC 1440
DB 1556 AACATAATGAAGAAGCTGAGCCCAACAGGAGGAGTGAACCAAGTTTAAAGTTTGAC 1615
QY 1441 TACAGCAGAGAGAGAACTACAGCACAATTAACACTGAATTTACAGATGAATGGAGC 1500
DB 1616 TACAGCAGAGAGAGAACTACAGCACAATTAACACTGAATTTACAGATGAATGGAGC 1675
QY 1501 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGGCTTGAACATCTGAGGGTCAGAGGAA 1560
DB 1676 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGGCTTGAACATCTGAGGGTCAGAGGAA 1735
QY 1561 TCTTTCTGTCACAGAGCTCAGTGCACCAACAGAGAGAGAGTGAACAAAGCTCT 1620
DB 1736 TCTTTCTGTCACAGAGCTCAGTGCACCAACAGAGAGAGAGTGAACAAAGCTCT 1795
QY 1621 GAAGATCATCAGAGGATGTGACAAATATACAGAGGAGTATCTGCAGAAAACCCAGTT 1680

Db 1796 GAAGATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAACCCAGTT 1855
 QY 1681 GAGAACCATATCAATATAACACCAATCAGATAAGTTTCACGCCCAAGCCATTGGATTCCAAC 1740
 Db 1856 GAGAACCATATCAATATAACACCAATCAGATAAGTTTCACGCCCAAGCCATTGGATTCCAAC 1915
 QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTCGCTCTGTGGGTTCCAGAAAGAAATCTGCT 1800
 Db 1916 TCAGGAGAAAGAAATGACCTCAATCTTGTCGCTCTGTGGGTTCCAGAAAGAAATCTGCT 1975
 QY 1801 TCATCTGAAAAAGCCAAAGAACCCAGAACTTCAGATCAGACTAGCAGTGTGCTTAC 1860
 Db 1976 TCATCTGAAAAAGCCAAAGAACCCAGAACTTCAGATCAGACTAGCAGTGTGCTTAC 2035
 QY 1861 AATGAAAAATACCAATCTCTGAGGCTCAGTTCCTCAAAAGAACCCAGTGTGCTTAC 1920
 Db 2036 AATGAAAAATACCAATCTCTGAGGCTCAGTTCCTCAAAAGAACCCAGTGTGCTTAC 2095
 QY 1921 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGCTTAC 1980
 Db 2096 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGCTTAC 2155
 QY 1981 CCAGTCTGTGTCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2040
 Db 2156 CCAGTCTGTGTCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2215
 QY 2041 GCCCGTATTCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2100
 Db 2216 GCCCGTATTCAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2275
 QY 2101 ACTTTGAACATTTAGAGGCGCTAGTAAATGTTTATAAGGCCATCCCACTCCAGG 2160
 Db 2276 ACTTTGAACATTTAGAGGCGCTAGTAAATGTTTATAAGGCCATCCCACTCCAGG 2335
 QY 2161 ACAATGATAAAGAGCCAAATTTCTGGGTCGTAATGTTTATAAGGCCATCCCACTCCAGG 2220
 Db 2336 ACAATGATAAAGAGCCAAATTTCTGGGTCGTAATGTTTATAAGGCCATCCCACTCCAGG 2395
 QY 2221 GGCACATTTTCATCTGGATCGGACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2280
 Db 2396 GGCACATTTTCATCTGGATCGGACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2455
 QY 2281 AATCATGTTGTTAACTGCTGAGCAGCAGATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2340
 Db 2456 AATCATGTTGTTAACTGCTGAGCAGCAGATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2515
 QY 2341 ATAGATTTAGATTAAGATCTGCTGAGCAGCAGATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2400
 Db 2516 ATAGATTTAGATTAAGATCTGCTGAGCAGCAGATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2575
 QY 2401 CTGCTGATGAAGTTTAACTCGAAACCACTCATGCTGGAAGAACTTAGAAACCACTT 2460
 Db 2576 CTGCTGATGAAGTTTAACTCGAAACCACTCATGCTGGAAGAACTTAGAAACCACTT 2635
 QY 2461 ACAGTTCAGGCTTTTCATCTGAGGATGTGCTTCACTTAATCATATCCGAGCTGAC 2520
 Db 2636 ACAGTTCAGGCTTTTCATCTGAGGATGTGCTTCACTTAATCATATCCGAGCTGAC 2695
 QY 2521 CGGTTGAGGCTGACAGATCAGAGGCTCTGCTCAAGAGATGAATGAGGATGAGGAA 2580
 Db 2696 CGGTTGAGGCTGACAGATCAGAGGCTCTGCTCAAGAGATGAATGAGGATGAGGAA 2755

RESULT 3

AX078303
 LOCUS
 DEFINITION Sequence 3022 bp DNA linear PAT 22-FEB-2001
 ACCESSION AX078303
 VERSION AX078303.1
 KEYWORDS human.
 SOURCE Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 3022)
 Hillman, J.L., Lal, P., Tang, Y.T., Yue, H., Au-Young, J., Bandman, O.,
 Azimzai, Y., Yang, J., Lu, D.A., Baughn, M.R., Patterson, C. and Shah, P.
 Cell cycle and proliferation proteins
 Patent: WO 0107471-A 107 01-FEB-2001;
 Incyte Genomics, Inc. (US)

FEATURES

Source

Location/Qualifiers

1..3022

/organism="Homo sapiens"

/db_xref="taxon:9606"

/note="Incyte ID No: 5682976CB1"

BASE COUNT 937 a 611 c 674 g 800 t

Query Match

Best Local Similarity 97.2%; Score 2508.4; DB 6; Length 3022;

Matches 2579; Conservative 0; Mismatches 1; Indels 60; Gaps 1;

QY 1 ATGTCCTCGGGTGGCTCTTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 60
 Db 119 ATGTCCTCGGGTGGCTCTTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 178
 QY 61 CTGAGGAGCCCGTCCCGGCTGCGGAGTCTACCTTGGGAGAGAGAAATTTATCCAAAGA 120
 Db 179 CTGAGGAGCCCGTCCCGGCTGCGGAGTCTACCTTGGGAGAGAGAAATTTATCCAAAGA 238
 QY 121 TTAACACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAACTCTGTGGAAT 180
 Db 239 TTAACACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAACTCTGTGGAAT 298
 QY 181 GACACTCGGAATATATTTATCTGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 240
 Db 299 GACACTCGGAATATATTTATCTGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 358
 QY 241 TACAGCAGAAAGGTTTGAACAACATTCGTTTCAGGACCCGAGCAACATATTTAGTGCA 300
 Db 359 TACAGCAGAAAGGTTTGAACAACATTCGTTTCAGGACCCGAGCAACATATTTAGTGCA 418
 QY 301 AACTCTTACCTTGTACAAATGATTAACAGATTTGTATCTCTCTGAGATGAGTAATA 360
 Db 419 AAGTTCTTACCTTGTACAAATGATTAACAGATTTGTATCTCTCTGAGATGAGTAATA 478
 QY 361 TTTTATACCAAGGTTGAGCAAGATGCAAGAACCAACAGACAAATGCAATTTACGTGTCAT 420
 Db 479 TTTTATACCAAGGTTGAGCAAGATGCAAGAACCAACAGACAAATGCAATTTACGTGTCAT 538
 QY 421 TATGNACTACTTATGAGATTTAGCTGTACCAATGACCTTACACCTTACACTTTCTCTTGT 480
 Db 539 TATGNACTACTTATGAGATTTAGCTGTACCAATGACCTTACACTTTCTCTTGT 598
 QY 481 GGTGAAGATGGAACCTGTAGGTTGTTGATACACCATCAAACTAGCTGCAAAAGAA 540
 Db 599 GGTGAAGATGGAACCTGTAGGTTGTTGATACACCATCAAACTAGCTGCAAAAGAA 658
 QY 541 CATTTGAAGATGATATTTTAACTCTGACGTGCTGCCACCTCTGCTCTGCTTATGTC 600
 Db 659 CATTTGAAGATGATATTTTAACTCTGACGTGCTGCCACCTCTGCTCTGCTTATGTC 718
 QY 601 CCACCAATACCATATTTACCTTCTCCCATCTTAATAAGTCTCTGACAGTGTGATATGAT 660
 Db 719 CCACCAATACCATATTTACCTTCTCCCATCTTAATAAGTCTCTGACAGTGTGATATGAT 778
 QY 661 CGGGAATCTGGGCAAGAGCTTACAGGAAATTTATGAGTCTGAGGAGTCTACTGGAATG 720
 Db 779 CGGGAATCTGGGCAAGAGCTTACAGGAAATTTATGAGTCTGAGGAGTCTACTGGAATG 838
 QY 721 GTTCCCGCTTTTATCTCTCCCATCTTAATAAGTCTCTGACAGTGTGATCTCTGCT 780
 Db 839 GTTCCCGCTTTTATCTCTCCCATCTTAATAAGTCTCTGACAGTGTGATCTCTGCT 898
 QY 781 TACAGTGAAGATGGTCAAGAGATTTCTGTTAGTTACTCTTTCAGATATATATCTTTT 840

Db	1979	TCATCTGAAAAGCCAAAGAACCCAGAAAACCTTCAGATCAGACTAGCACTGAGAGTGCCTACC	2038
QY	1861	AATGAAATTAACACCAATCTCAGCTCAGTTCCTCAACACAGAACCCACCTGGGCTTTCAGCT	1920
Db	2039	AATGAAATTAACACCAATCTCAGCTCAGTTCCTCAACACAGAACCCACCTGGGCTTTCAGCT	2098
QY	1921	CATGGAAGAACATCCACACGAGGACTCTGCTCTTCAGGACACAGATGACAGTGAATGATAC	1980
Db	2099	CATGGAAGAACATCCACACGAGGACTCTGCTCTTCAGGACACAGATGACAGTGAATGATAC	2158
QY	1981	CCAGTCTGTATCCAGGTCGAAGGTATCCAGCAGGACCTGGTGATAGACGCTCTGCTGTT	2040
Db	2159	CCAGTCTGTATCCAGGTCGAAGGTATCCAGCAGGACCTGGTGATAGACGCTCTGCTGTT	2218
QY	2041	GCCGCTATTCAGGAGTCTTTCAGACGAGAGAAAAGAAAGAAAGAAATGGAAGAAATTTGGAT	2100
Db	2219	GCCGCTATTCAGGAGTCTTTCAGACGAGAGAAAAGAAAGAAAGAAATGGAAGAAATTTGGAT	2278
QY	2101	ACTTTGAACATTAAGAAGCCGCTAGTAAAATGGTTTTATAAGGCCATCGCAACTCCAGG	2160
Db	2279	ACTTTGAACATTAAGAAGCCGCTAGTAAAATGGTTTTATAAGGCCATCGCAACTCCAGG	2338
QY	2161	ACAATGATAAAGAAGCCAAATTCCTGGGGTGCTAACTTTGTAAATGAGTGGTTCGACTGT	2220
Db	2339	ACAATGATAAAGAAGCCAAATTCCTGGGGTGCTAACTTTGTAAATGAGTGGTTCGACTGT	2398
QY	2221	GGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGATCTTCTCGGAAGCTGAT	2280
Db	2399	GGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGATCTTCTCGGAAGCTGAT	2458
QY	2281	AATCATGTGGTAAACTGCTCGAGCCACATCGGTTTGACCCAAATTTTAGCCTCATCTGGC	2340
Db	2459	AATCATGTGGTAAACTGCTCGAGCCACATCGGTTTGACCCAAATTTTAGCCTCATCTGGC	2518
QY	2341	ATAGATTATGACATAAAGATCTGGTCACCATTTAGAAGAGTCAAGGATTTTAAACCGAAA	2400
Db	2519	ATAGATTATGACATAAAGATCTGGTCACCATTTAGAAGAGTCAAGGATTTTAAACCGAAA	2578
QY	2401	CTTGCTGTGATGAATTAAACTCGAAACGAACCTCATGCTGGAAGAACTAGAAACACCAATT	2460
Db	2579	CTTGCTGTGATGAATTAAACTCGAAACGAACCTCATGCTGGAAGAACTAGAAACACCAATT	2638
QY	2461	ACAGTTTCCAGCCCTTTTCATGTTGAGGATGTTGGCTTTCATTAACTATCTCGAGCTGAC	2520
Db	2639	ACAGTTTCCAGCCCTTTTCATGTTGAGGATGTTGGCTTTCATTAACTATCTCGAGCTGAC	2698
QY	2521	CGGTTGAGGGTGACAGATCAGAGGCTCTGGTCAAGAGAAATGAAAATGAGGATGAGGAA	2580
Db	2699	CGGTTGAGGGTGACAGATCAGAGGCTCTGGTCAAGAGAAATGAAAATGAGGATGAGGAA	2758

RESULT 4				PAT 22-FEB-2001
AX076995				
LOCUS		3420 bp	DNA	linear
DEFINITION	Sequence 3 from Patent WO0107616.			
ACCESSION	AX076995			
VERSION	AX076995.1	GI:13121648		
KEYWORDS	.			
SOURCE	human.			
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 3420)			
AUTHORS	el Shami,A.S., Menon,S.N. and French,C.K.			
TITLE	Poly nucleotide encoding autoantigens associated with endometriosis			
JOURNAL	Patent: WO 0107616-A 3 01-FEB-2001;			
	DIAGNOSTIC PRODUCTS CORPORATION (US)			
FEATURES	Location/Qualifiers			
source	1..3420 /organism="Homo sapiens" /db_xref="taxon:9606" 176..2989 /note="unnamed protein product"			
CDS				

```

/codon_start=1
/protein_id="CAC32385.1"
/db_xref="GI:13121649"
/translation="NSRGSPHLLWVRKFLGLDPSRLSRVLRGRFRIORLKL
ATLVNDGCVNTICWNDGEGYLLSGSDDTKLVI SNPYSRKVLTTIRSHRANISAKF
LPCTNDKIOVSCSDGVYIFNYNBOADETNRQCOFTCHYGTYYEIMVNDPYFLSC
GEDGVNFDRIKTSCTKEDKDDILINCRRAATVAICPPYLYAVGCSDSVRI
YDRMLGTRATAGTGGTTCGVNARFIPSHLNKRSVTSYSLYSEDGELLVSYSDY
LYLPDKDARELKTSAERREBELPQPVKRLRLCDMSDITGPRAPRSEERDGE
QSPNVSLMDMLSRWFEEASVAOSNRGRGRSRRPGTSGSDISTLPVPSDDL
EVSTAMEVDTPAQOFLQPTSSSTMSQAUSTSPSPHSTPLSSPSDEQRQSVBA
SGHITHQSEFLRGPETALLRKLQQLRLAKAQROQELAAHTQQPSTSDOSHHG
SSQDPHSDSPSSVYNKOLGMSLSDQNNNEKLSKPKGTPVLSLHYSTRGTTS
TKFNFDWSSIASSRGSGSESESVPOSSVOPGEGDSGETKAPESSEYV
KEYDEGSAENPVNHHITQDSKFTAKPLDSNSGRNDLNLDRSGVPRESSEKA
KEPSTQSTESGDSAVARIQEFRRRKRKEMEELDTLNLRLPYKMYKGRHNSRT
IPNARYRAGPGRSARVARIOEFRRRKRKEMEELDTLNLRLPYKMYKGRHNSRT
MIKANFWPFLVMTGSECHIFIDWHTAHLMLLEADNHVYVNCIQLPFPFPIILASS
GIDYDKINSPLESIRFNKLADDEVITRNELMLEETRNITVPASFMLRLASLNIH
RADRLGDRSGEGQENENEDE"
BASE COUNT 1036 a 723 c 778 g 883 t
ORIGIN

```

```

Query Match          90.4%; Score 2332.6; DB 6; Length 3420;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 2576; Conservative 0; Mismatches 4; Indels 231; Gaps 1;

QY 1 ATGTCGGGGTGGCCCTACCCACACCTGTTGTTGGAGCTGAGGAAAAGTCCCTCGGG 60
DB 176 ATGTCGGGGTGGCCCTACCCACACCTGTTGTTGGAGCTGAGGAAAAGTCCCTCGGG 235
QY 61 CTGGAGGACCCGCGCGTGGGAGTGCCTACCTGGGAAGAAGAGATTTATCCAAAGA 120
DB 236 CTGGAGGACCCGCGCGTGGGAGTGCCTACCTGGGAAGAAGAGATTTATCCAAAGA 295
QY 121 TTAACACTTGAAGCAACCCCTTAATGTCATGATGTTGTTGTTAATACAACTGTTGGAAT 180
DB 296 TTAACACTTGAAGCAACCCCTTAATGTCATGATGTTGTTGTTAATACAACTGTTGGAAT 355
QY 181 GACACTGGAGATATATTTATCTGGCTGACATGACACCAATAGTAATAGTAATCTCT 240
DB 356 GACACTGGAGATATATTTATCTGGCTGACATGACACCAATAGTAATAGTAATCTCT 415
QY 241 TACAGCAGAAAGTTTGGACAACTTCTGTTGAGGACCCAGCAGCAACATATTTAGTGCA 300
DB 416 TACAGCAGAAAGTTTGGACAACTTCTGTTGAGGACCCAGCAGCAACATATTTAGTGCA 475
QY 301 AAGTCTTACCTGTACAAATGATAACAGATTGTATCTGCTGAGATGGAGTAATA 360
DB 476 AAGTCTTACCTGTACAAATGATAACAGATTGTATCTGCTGAGATGGAGTAATA 535
QY 361 TTTTATACCAACCTTGAGCAAGATGCAGAAACCAACAGCAATGCCAATTTACGTGTCAT 420
DB 536 TTTTATACCAACCTTGAGCAAGATGCAGAAACCAACAGCAATGCCAATTTACGTGTCAT 595
QY 421 TATGGAACCTACTTATGAGATATGACTGTACCAATGACCCCTTACACTTTTCTCTTGT 480
DB 596 TATGGAACCTACTTATGAGATATGACTGTACCAATGACCCCTTACACTTTTCTCTTGT 655
QY 481 GGTGAGATGGAACCTGTTAGTGGTGTGATACACGATCAAACTAGCTGCACAAAAGAA 540
DB 656 GGTGAGATGGAACCTGTTAGTGGTGTGATACACGATCAAACTAGCTGCACAAAAGAA 715
QY 541 GATTGTAAGATGATATTTTAAATTAACGTGACGTGCTGCCAGCTGTGTTGCTATTGTC 600
DB 716 GATTGTAAGATGATATTTTAAATTAACGTGACGTGCTGCCAGCTGTGTTGCTATTGTC 775
QY 601 CCACCAATACCATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
DB 776 CCACCAATACCATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 835
QY 661 CGGGGAATGCTGGGCACAGAGCTACAGGGAATTTATCGAGTGGAGGACTACTTGGATG 720
DB 835 CGGGGAATGCTGGGCACAGAGCTACAGGGAATTTATCGAGTGGAGGACTACTTGGATG

```

```

DB 836 CGGGGAATGCTGGGCACAGAGCTACAGGGAATTTATCGAGTGGAGGACTACTTGGATG 895
QY 721 GTTGGCCCTTTTATTTCCCTTCCCATCTTAATAATAGTCTCTCAGAGTGCACATCTCTGTGT 780
DB 896 GTTGGCCCTTTTATTTCCCTTCCCATCTTAATAATAGTCTCTCAGAGTGCACATCTCTGTGT 955
QY 781 TACAGTGAAGATGCTCAAGAGATTCTGTTAGTTACTTCTTCCAGATTACATATATCTTTTT 840
DB 956 TACAGTGAAGATGCTCAAGAGATTCTGTTAGTTACTTCTTCCAGATTACATATATCTTTTT 1015
QY 841 GACCCGAAGATGATACAGCACGAGAACTTAAACTCTTCTCGGGAAGAGAGAGAGAA 900
DB 1016 GACCCGAAGATGATACAGCACGAGAACTTAAACTCTTCTCGGGAAGAGAGAGAGAA 1075
QY 901 GAGTTCGGCACAACCAACAGTTAAGCGTTTTCAGACTTTCGTTGGTATTTGTTGTCAGATCTGGA 960
DB 1076 GAGTTCGGCACAACCAACAGTTAAGCGTTTTCAGACTTTCGTTGGTATTTGTTGTCAGATCTGGA 1135
QY 961 CCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGGAGAGCAGAGTCCCAATGTGCA 1020
DB 1136 CCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGGAGAGCAGAGTCCCAATGTGCA 1195
QY 1021 TTGATCCAGAGATGCTGTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1080
DB 1196 TTGATCCAGAGATGCTGTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1255
QY 1081 CAAAGCAATAGAGGACGAGGAGATCTCGCCAGAGAGTGGAAAGCAAGTCAATCAGATATT 1140
DB 1256 CAAAGCAATAGAGGACGAGGAGATCTCGCCAGAGAGTGGAAAGCAAGTCAATCAGATATT 1315
QY 1141 TCAACTCTTCTAGGTCCTCCATCAAGTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
DB 1316 TCAACTCTTCTAGGTCCTCCATCAAGTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1375
QY 1201 GTAGATCTCCAGCTGAACAAATTTCTTCAAGCTTCTACATCTCTACATCTCTCAGCTCAG 1260
DB 1376 GTAGATCTCCAGCTGAACAAATTTCTTCAAGCTTCTACATCTCTCAGCTCAG 1435
QY 1261 GCTCATTTCGACATCTCTCCACAGAAAGCCCTCAATCTACTCTTCTTCTTCTTCTTCTTCTTCT 1320
DB 1436 GCTCATTTCGACATCTCTCCACAGAAAGCCCTCAATCTACTCTTCTTCTTCTTCTTCTTCTTCT 1495
QY 1321 GACAGTGAACAAAGGAGCTGTTGAGGATCTGGACACACACACATCATCAGTCT --- 1377
DB 1496 GATAGTGAACAAAGGAGCTGTTGAGGATCTGGACACACACACATCATCAGTCT --- 1555
QY 1378 --- 1377
DB 1556 TTTTAAAGGGCGCTGAGATAGCTTTGCTTCTGTAAGCGCTGCAACAACTGAGGCTTAAG 1615
QY 1378 --- 1377
DB 1616 AAGGCTGAGCAGCAGAGCAGCAGAGCTAGCTGCACATACCCAGCAACAGGCTTCCACT 1675
QY 1378 --- 1377
DB 1676 TCTGATCAGTCTTCTCAGAGGGCTCTTTCAGGACCCCTCATGCTTCAGATTCCTCTCT 1735
QY 1378 --- 1377
DB 1736 TCTGTGTTAAACAAAGCTCGGATCCATGCTGCTGACGAGCAACAGGATAACAATAAT 1795
QY 1390 GAAAGCTGAGCCCAACCAAGGACAGGAGTGAACCAAGTGTAAAGTTTGAAGTGTACAGCA 1449
DB 1796 GAAAGCTGAGCCCAACCAAGGACAGGAGTGAACCAAGTGTAAAGTGTACAGCA 1855
QY 1450 GAGGAACTACTAAGCAACATATAAAGTGAAGTGTACAGATGAATGGAGCAGTATAGCA 1509
DB 1856 GAAAGCAACTACAGCAACATATAAAGTGAAGTGTAAAGTGTACAGATGAATGGAGCAGTATAGCA 1915
QY 1510 TCAAGTCTTAGAGGATTTGGGAGCCATTCGAATCTGAGGCTCAGGAGGAACTTCTTCTGTC 1569
DB 1916 TCAAGTCTTAGAGGATTTGGGAGCCATTCGAATCTGAGGCTCAGGAGGAACTTCTTCTGTC 1975

```

Sequence 3 from Patent EP1106690.	
DEFINITION	AX113968
ACCESSION	AX113968.1
VERSION	GI:13940142
KEYWORDS	human.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 3420)
TITLE	el Shami,A.S., Menon,S.N. and French,C.K.
JOURNAL	polynucleotide encoding autoantigens associated with endometriosis
DIAGNOSTIC PRODUCTS CORPORATION (US)	Patent: EP 1106690-A 3 13-JUN-2001.
FEATURES	Location/Qualifiers
source	1..3420
CDS	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	176..2989
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAC37868.1"
	/db_xref="GI:13940143"
	/translation="MSRGGSPHLLWVKKRFLGLEDPRLRSRYLGRREFIQRILKE
	ATLVNHDGCVNTICWNRDGEYILSGDDIKLVISNPYSRKYLTIRSRHRANIFSAKF
	LPTNDKQIVSCSGDGVTFVTNVEQDAETNOCQFCHYGTIYETIMVNDPYFLSC
	GEDGVRFEDTRIKTSCTKEDCKDDILNCRRAATSVATCPIPIYLAAGSDSVRI
	YDRMLGTATCNVAGRGTTGMVRFELPPVRLRLRGDWSIDTGPRAPESEKRDGE
	IYLFDPKDTARELKTPSAEERRELPVRLRLRGDWSIDTGPRAPESEKRDGE
	QSPNSLMQMSDMSLRWFEAEVAQNRGRSRPRGTTSSDSTLPTVPSPDL
	EYSETEMEVDTPAEQFLQSTSTWASQAQHSSTSPSPHSTPLSSPSQSSHEG
	SGHTTHQSEFLRGPELALLRLKQQLRLKKAQOORQBLAAHTQOQPSSTQSSHEG
	SSODPHASDSSSVNKKQJLMSLDLQDDNNNEKLSPKPTGEPVLSLHYSTEGTTS
	TKLNTDEWSSVASSRIGISCKSGQSEFVPOSSVOPPEGSETKAPESSEDA
	TKYQGVSAENPVENINITQSDKFTKPLDSNGERNDLNDLRSCGVPESSEDA
	KEPESDOTSTESATNENINTPQFQTEATGSAHEETSDRLSDLODSDDDPVL
	IPGARYAGPDARSVAVARIOEFFFFRRRKEKEMEELDTLNRPLMKYKGRNRT
	MIKRANFWAGNFVMTGSECGHIFTDWHTAAHLMLLEADNHVNCLOPHFPDLASS
	GIDYDIKWSPLESRIFNEKLADVITRNLMEETRTITVPSFMRMLASLAHI
	RADRLGDRSGSGQENENEDER"
BASE COUNT	1036 a 723 c 778 g 883 t
ORIGIN	
Query Match	90.4%; Score 2332.6; DB 6; Length 3420;
Best Local Similarity	91.6%; Pred. No. 0;
Matches 2576; Conservative	0; Mismatches 4; Indels 231; Gaps 1;
QY 1	ATGCTCGGGGTGGCTCTACCCACACCTGTTGTGGAGCTGAGGAAAGGTCCTCGGG 60
DB 176	ATGCTCGGGGTGGCTCTACCCACACCTGTTGTGGAGCTGAGGAAAGGTCCTCGGG 235
QY 61	CTGAGGACCCGTCGGGCTGCGAGTCTGCTACCTGGGAGAGAGAAATTTATCCAAAGA 120
DB 236	CTGAGGACCCGTCGGGCTGCGAGTCTGCTACCTGGGAGAGAGAAATTTATCCAAAGA 295
QY 121	TTAAACCTTGAAGCAACCTTAATGTCATGATGTTGTGTTAATAACAATCTGTGGAAT 355
DB 296	TTAAACCTTGAAGCAACCTTAATGTCATGATGTTGTGTTAATAACAATCTGTGGAAT 355
QY 181	GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATAGTAATAGTAATCTCT 240
DB 356	GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATAGTAATAGTAATCTCT 415
QY 241	TACACGAGAAAGGTTTTCACACCAATTCGTTGAGGACCCGAGCAACAAATTTAGTGCA 300
DB 416	TACACGAGAAAGGTTTTCACACCAATTCGTTGAGGACCCGAGCAACAAATTTAGTGCA 475
QY 301	AAGTTCTTACCTTGTACAAATGATAAAGATTTGATCTCTCTGGAGATGGAGTAATA 360
DB 476	AAGTTCTTACCTTGTACAAATGATAAAGATTTGATCTCTCTGGAGATGGAGTAATA 535
QY 361	TTTTATACCAACGTTGACCAAGATGAGAAACCAACAGCAATGCAATTTACGTGTCTAT 420
DB 536	TTTTATACCAACGTTGACCAAGATGAGAAACCAACAGCAATGCAATTTACGTGTCTAT 595

QY 421 TATGGAACACTTATGAGATTATGACTGTACCAATGACCCTTACACTTTTCTCTCTGT 480
Db 596 TATGGAACACTTATGAGATTATGACTGTACCAATGACCCTTACACTTTTCTCTCTGT 655
QY 481 GGTGAAGATGGAACCTTATGAGTTTGTATACAGCATCAAACTAGCTGCACAAAGAA 540
Db 656 GGTGAAGATGGAACCTTATGAGTTTGTATACAGCATCAAACTAGCTGCACAAAGAA 715
QY 541 GATTGTAAGATGATATTTTAACTGTGACGTGCTGCCAGTGTGTGCTATTGTC 600
Db 716 GATTGTAAGATGATATTTTAACTGTGACGTGCTGCCAGTGTGTGCTATTGTC 775
QY 601 CCACCAATACCATATACCTTGTGTTGTTGTGACAGCTCAGTACCAATATATGAT 660
Db 776 CCACCAATACCATATACCTTGTGTTGTTGTGACAGCTCAGTACCAATATATGAT 835
QY 661 CGCGAATGCTGGGCAACAGAGCTACAGGGAATATGACAGTGCAGGGAATGGAATG 720
Db 836 CGCGAATGCTGGGCAACAGAGCTACAGGGAATATGACAGTGCAGGGAATGGAATG 895
QY 721 GTTGCCCGTTTATCTCTCCATCTTATATATAGTCTGTCAGAGTGCAGATCTCTGTGT 780
Db 896 GTTGCCCGTTTATCTCTCCATCTTATATATAGTCTGTCAGAGTGCAGATCTCTGTGT 955
QY 781 TACAGTGAAGATGCTCAAGAGATCTCTGTTAGTACTCTTACAGATTACATATATCTTTT 840
Db 956 TACAGTGAAGATGCTCAAGAGATCTCTGTTAGTACTCTTACAGATTACATATATCTTTT 1015
QY 841 GACCCGAAGATGATACAGCAGAGACTTAAACTCTCTGCGGAAGAGAGAGAA 900
Db 1016 GACCCGAAGATGATACAGCAGAGACTTAAACTCTCTGCGGAAGAGAGAGAA 1075
QY 901 GAGTTGGACAAACACAGTTAGCGTTTGTGAGACTCTGTTGATTTGTTGCTCAGATCTGGA 960
Db 1076 GAGTTGGACAAACACAGTTAGCGTTTGTGAGACTCTGTTGATTTGTTGCTCAGATCTGGA 1135
QY 961 CCAGAGCAAGCGCGAGAGTGAACAGAGAGATGAGAGAGAGTCCCAATGTGCA 1020
Db 1136 CCAGAGCAAGCGCGAGAGTGAACAGAGAGATGAGAGAGAGTCCCAATGTGCA 1195
QY 1021 TTGATGCAGAGATGCTCATATGTTATCAAGTGTGTTGAAGAGCAAGTCAATCATGAT 1080
Db 1196 TTGATGCAGAGATGCTCATATGTTATCAAGTGTGTTGAAGAGCAAGTCAATCATGAT 1255
QY 1081 CAAGCAATAGAGGAGAGAGATCTCGACCCAGAGTGAACAGTCAATCATGAT 1140
Db 1256 CAAGCAATAGAGGAGAGATCTCGACCCAGAGTGAACAGTCAATCATGAT 1315
QY 1141 TCAACTCTTCTACGTTCCCATCAAGTCTGATTTGGAAGTGAAGTGAACAGTCAATCATGAT 1200
Db 1316 TCAACTCTTCTACGTTCCCATCAAGTCTGATTTGGAAGTGAAGTGAACAGTCAATCATGAT 1375
QY 1201 GTAGTACTCCAGTGAACAAATTTCTTACGCTTCTACATCTCTCAATGTGAGTCAAG 1260
Db 1376 GTAGTACTCCAGTGAACAAATTTCTTACGCTTCTACATCTCTCAATGTGAGTCAAG 1435
QY 1261 GCTCATTCGACATCTCTCCAGAGAGCCCTCATCTCTCTTGTGATCTCTCA 1320
Db 1436 GCTCATTCGACATCTCTCCAGAGAGCCCTCATCTCTCTTGTGATCTCTCA 1495
QY 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGACACACACATCATCAGTCT 1377
Db 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGACACACACATCATCAGTCT 1555
QY 1378 ----- 1377
Db 1556 TTTTAAAGGGGCTTATGATAGCTTTGCTTGTAGAGCCCTGCAACAACTGAGGCTTAAG 1615
QY 1378 ----- 1377
Db 1616 AAGGCTGAGCAGAGGAGAGTGTGACATACCCAGAGAGGCTTCCACT 1675

QY 1378 ----- 1377
Db 1676 TCTGATCAGTCTTCTCATGAGGCTCTTACAGGACCCTCATGCTTCAGATTCTCTTCT 1735
QY 1378 ----- 1377
Db 1736 TCTGTTGTTAAACACAGCTCGGATCGATCTCTTACCTTGCAGCAGCAACAGGATACATAAT 1389
QY 1390 GAAAGCTGAGCCCAACACAGGAGAGTGAACCACTTTTAACTTGCACATACATAAT 1795
Db 1796 GAAAGCTGAGCCCAACACAGGAGAGTGAACCACTTTTAACTTGCACATACATAAT 1449
QY 1450 GAAGGAACTTACAGCAATAAACTTAACTTTACAGATGAATGAGCAGTATAGCA 1855
Db 1856 GAAGGAACTTACAGCAATAAACTTAACTTTACAGATGAATGAGCAGTATAGCA 1509
QY 1510 TCAAGTTCTAGAGGAATTTGGAGCCATTTGCAAACTCTGAGGCTCAGGAGGAATCTTCTGTC 1915
Db 1916 TCAAGTTCTAGAGGAATTTGGAGCCATTTGCAAACTCTGAGGCTCAGGAGGAATCTTCTGTC 1569
QY 1570 CCACAGAGCTCAGTGCACCAACAGGAGAGTGAACCACTTTTAACTTGCACATACATAAT 1975
Db 1976 CCACAGAGCTCAGTGCACCAACAGGAGAGTGAACCACTTTTAACTTGCACATACATAAT 1629
QY 1630 TCAGAGGATGTGACAAATATCAGGAAGAGTATCTGCAGAAACCCAGTTCGAGAACCAT 2035
Db 2036 TCAGAGGATGTGACAAATATCAGGAAGAGTATCTGCAGAAACCCAGTTCGAGAACCAT 1689
QY 1690 ATCAATATACACAACTCAGATAAGTTTCACAGCCCAAGCCATTTGATTTCCAACTCAGGAGAA 2095
Db 2096 ATCAATATACACAACTCAGATAAGTTTCACAGCCCAAGCCATTTGATTTCCAACTCAGGAGAA 1749
QY 1750 AGAATGACCTCAATCTTGTATGCTCTTGTGGGTTTCCAGAAAGATCTGCTTCACTGAA 2155
Db 2156 AGAATGACCTCAATCTTGTATGCTCTTGTGGGTTTCCAGAAAGATCTGCTTCACTGAA 1809
QY 1810 AAAGCCAGGAGACCAAACTTACAGATCAGACTAGCAGTACAGTGCAGTGCATCAATGAAAT 2215
Db 2216 AAAGCCAGGAGACCAAACTTACAGATCAGACTAGCAGTACAGTGCATCAATGAAAT 1869
QY 1870 AACCAATCTTGCAGCTTCACTTCCAAACAGAGCCACTGGGCTTTCAGCTCATGAGAA 2275
Db 2276 AACCAATCTTGCAGCTTCACTTCCAAACAGAGCCACTGGGCTTTCAGCTCATGAGAA 1929
QY 1930 ACATCCACAGGAGTCTGCTTTCAGGACACAGATGACAGTGCAGTGCATGAGTGCCTG 2335
Db 2336 ACATCCACAGGAGTCTGCTTTCAGGACACAGATGACAGTGCAGTGCATGAGTGCCTG 1989
QY 1990 ATCCAGGTCGAAGGTATCAGCAGGAGCTGTTGATAGAGCTCTCTGTTGCCGCTATT 2395
Db 2396 ATCCAGGTCGAAGGTATCAGCAGGAGCTGTTGATAGAGCTCTCTGTTGCCGCTATT 2049
QY 2050 CAGGAGTCTTTCAGAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2455
Db 2456 CAGGAGTCTTTCAGAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2109
QY 2110 ATTAGAAGCCGCTAGTAAATGTTTATAGAGCCATCGCAACTCCAGGACATGATA 2515
Db 2516 ATTAGAAGCCGCTAGTAAATGTTTATAGAGCCATCGCAACTCCAGGACATGATA 2169
QY 2170 AAAGAAGCAATTTCTGGGCTGCTAACTTTGATAGTGTGTTCTGACTGTGGCCACTT 2575
Db 2576 AAAGAAGCAATTTCTGGGCTGCTAACTTTGATAGTGTGTTCTGACTGTGGCCACTT 2229
QY 2230 TTCACTTGGATCGGACACTGCTGAGCATTTGATGCTTCTGAGAGTGCATGATAATGATG 2635
Db 2636 TTCACTTGGATCGGACACTGCTGAGCATTTGATGCTTCTGAGAGTGCATGATAATGATG 2289
QY 2290 GTAACTGCTGCAGCAGATCTGTTGACCAATTTAGCCCTCATCTGGCATAGATTAT 2695
Db 2696 GTAACTGCTGCAGCAGATCTGTTGACCAATTTAGCCCTCATCTGGCATAGATTAT 2349
QY 2350 GACATAAAGATCTGGTCAACCATTTAGAAGAGTCAAGGATTTTAAACGAAACTTCTGCTGAT 2755

us-09-781-693a-1.rge

Wed Mar 26 09:38:48 2003

Db	436	CAGTACGAATATATGATCGCGGAATGCTGGGCACAGAGCTACAGGGAATATGCGAGTC	495
Qy	704	GAGGACTACTGGAATGGTTGCCGTTTATCTCCCATCTTATATATATAGTCTGCGA	763
Db	496	GAGGACTACTGGAATGGTTG-CCGTTTATCTCCCATCTTATATATAGTCTGCGA	554
Qy	764	GAGTACATCTCTGTTTACAGTGAAGTGTCAAGAGATCTCTGTTAGTCTCTTCA	823
Db	555	GAGTACATCTCTGTTTACAGTGAAGTGTCAAGAGATCTCTGTTAGTCTCTTCA	614
Qy	824	ATTACATATATCTTTTACCCGGAAGATGATACAGCAGCAGAGACTTTAAACCTCT	883
Db	615	ATTACATATATCTTTTACCCGGAAGATGATACAGCAGCAGAGACTTTAAACCTCT	674
Qy	884	CGGAGAGAGAGAGAGAGTGTGCGACACACACCAAGTTAAGCGTTTGGAGTCTG	943
Db	675	CGGAGAGAGAGAGAGAGTGTGCGACACACCAAGTTAAGCGTTTGGAGTCTG	734
Qy	944	ATTGGTCAGATCTGACCCAGCAGCAAGCGCGAGAGTGAACAGAGATCGACCGAG	1003
Db	735	ATTGGTCAGATCTGACCCAGCAGCAAGCGCGAGAGTGAACAGAGATCGACCGAG	794
Qy	1004	AGAGTCCCAATGTGTCTATGATGAGAGAGTGTCTGATATGTTATCAAGATGTT	1063
Db	795	AGAGTCCCAATGTGTCTATGATGAGAGAGTGTCTGATATGTTATCAAGATGTT	854
Qy	1064	AAGCAATGAGTGTGCAACAAAGATAGAGAGAGAGAGTGTCTGATATGTTATCA	1123
Db	855	AAGCAATGAGTGTGCAACAAAGATAGAGAGAGAGAGTGTCTGATATGTTATCA	914
Qy	1124	CAAGTCAATCAGATATTTCAACTCTTCTGCGTCCCATCAAGTCTCTGATATGTT	1183
Db	915	CAAGTCAATCAGATATTTCAACTCTTCTGCGTCCCATCAAGTCTCTGATATGTT	974
Qy	1184	GTGAACATGCAATGAGTGTGATATCTCCAGCTGAACAAATTTCTCAGCCTCT	1243
Db	975	GTGAACATGCAATGAGTGTGATATCTCCAGCTGAACAAATTTCTCAGCCTCT	1034
Qy	1244	CTACATATGTCAGTCTGAGTCTATTCGACATCATCTCCACAGAAAGCCCTCAT	1303
Db	1035	CTACATATGTCAGTCTGAGTCTATTCGACATCATCTCCACAGAAAGCCCTCAT	1094
Qy	1304	CTTTGCTATCTTCTCCAGACAGTGAACAAAGCAGTCTGTTGAGGCATCTGAC	1363
Db	1095	CTTTGCTATCTTCTCCAGACAGTGAACAAAGCAGTCTGTTGAGGCATCTGAC	1154
Qy	1364	CACATATCAGTCTGATTAACAATAATGAAGTGTGAGC-----1401	1214
Db	1155	CACATATCAGTCTGATTAACAATAATGAAGTGTGAGC-----1401	1214
Qy	1402	-----CCCAACCCAGGACAGAGTGAACCCAG 1426	1274
Db	1215	TTGAGGACCAACAGGATACCAATATGAAGTGAAGGAGTGAACCCAG 1486	1426
Qy	1427	TTTTAAGTTTGCATCTACAGCAGCAAGAGAACTACACGACAAATAAAGTGA	1334
Db	1275	TTTTAAGTTTGCATCTACAGCAGCAAGAGAACTACACGACAAATAAAGTGA	1346
Qy	1487	CAGATGAATGAGCAGATATAGCATCAAGTCTAGAGGAATTTGGAGCCATTG	1394
Db	1335	CAGATGAATGAGCAGATATAGCATCAAGTCTAGAGGAATTTGGAGCCATTG	1606
Qy	1547	AGGGTCAGAGGAGTCTTTTCTGTCACAGAGCTCAGTGCACCCAGAGAGCAG	1454
Db	1395	AGGGTCAGAGGAGTCTTTTCTGTCACAGAGCTCAGTGCACCCAGAGAGCAG	1666
Qy	1607	AAACAAAGCTCTGGAAGATATCAGAGATGTCAGAAATATCAGGAAGAGTAT	1514
Db	1455	AAACAAAGCTCTGGAAGATATCAGAGATGTCAGAAATATCAGGAAGAGTAT	1726
Qy	1667	CAGAAACCCAGTGTGAGAACCATATCAATATAACAAATCAGATAGTTCAG	1574
Db	1515	CAGAAACCCAGTGTGAGAACCATATCAATATAACAAATCAGATAGTTCAG	1574

Db	2756	GACATAAAGATCTGCTCACCATCTAGAGAGTCAAGATTTTAAACGAAACTTGTGAT	2815
Qy	2410	GAGTATTAACCTCGAAGCACTCATCTGGAAGAACTAGAAACACCATACAGTTCCA	2469
Db	2816	GAGTATTAACCTCGAAGCACTCATCTGGAAGAACTAGAAACACCATACAGTTCCA	2875
Qy	2470	GCCTCTTTCATGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGACCGGTTGG	2529
Db	2876	GCCTCTTTCATGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGACCGGTTGG	2935
Qy	2530	GGTGACAGATCAGAGAGTGTGCTGCAAGAGAGATGAAATGAGAGTGAAGAA	2580
Db	2936	GGTGACAGATCAGAGAGTGTGCTGCAAGAGAGATGAAATGAGAGTGAAGAA	2986
RESULT 6			
AX060956	2448 bp	DNA	linear PAT 22-JAN-2001
LOCUS	Sequence 8 from Patent WO0100874.		
DEFINITION	AX060956		
ACCESSION	AX060956.1	GI:12406330	
VERSION			
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1. (bases 1 to 2448)		
JOURNAL	Sabin, V., Tureci, O. and Pfreundschuh, M.		
FEATURES	Cancer associated antigens and uses therefor		
	Patent: WO 0100874-A 8 04-JAN-2001;		
	LUDWIG INSTITUTE FOR CANCER RESEARCH (US)		
	Location/Qualifiers		
	1..2448		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	778 a 513 c 551 g 606 t		
ORIGIN			
Query Match	82.9%; Score 2140; DB 6; Length 2448;		
Best Local Similarity	95.4%; Pred. No. 0;		
Matches 2303; Conservative	0; Mismatches 45; Indels 56; Gaps 7;		
Qy	228	AATTAGTAACTCTTACAGCAGAAAGTTTGGACAAATTCGTTACAGGACCGAGCAAA	287
Db	16	AGTAAGGAGAGGTTTCTCATCAGAGGTTTGGACAAATTCGTTACAGGACCGAGCAAA	75
Qy	288	CATATTTAGTGCAAGTCTTACCTGTACAAATGATTAACAGATTTGATCTGCTCTGG	347
Db	76	CATATTTAGTGCAAGTCTTACCTGTACAAATGATTAACAGATTTGATCTGCTCTGG	135
Qy	348	AGATGAGTATATTTTATACCAAGTTGAGCAAGATGCAAGAACTACCAACAGCAATGCCA	407
Db	136	AGATGAGTATATTTTATACCAAGTTGAGCAAGATGCAAGAACTACCAACAGCAATGCCA	195
Qy	408	ATTACGCTGCTATATGGAACCTATTTAGATATGATGCTGACCTGACCTTACAC	467
Db	196	ATTACGCTGCTATATGGAACCTATTTAGATATGATGCTGACCTGACCTTACAC	255
Qy	468	TTTTCTCTCTTGTGGTGAAGA-TGGAACTGTTAGTGGTTGATACAGCATCAAACTA	526
Db	256	TTTTCTCTCTTGTGGTGAAGA-TGGAACTGTTAGTGGTTGATACAGCATCAAACTA	315
Qy	527	GCTGCACAAAGAGATGTTAAGATGATATTTAATTAACCTGCTGCTGCTGCTGCTG	586
Db	316	GCTGCACAAAGAGATGTTAAGATGATATTTAATTAACCTGCTGCTGCTGCTGCTG	375
Qy	587	CTGTTT--GCTATTTGCCCAACATATACATATTA-CCTTGTGTTGTTGTTGTTGCTG	643
Db	376	CTGTTT--GCTATTTGCCCAACATATACATATTA-CCTTGTGTTGTTGTTGTTGCTG	435
Qy	644	CAGTACGAATATATGATCGCGAATGCTGGGCACAGAGCTACAGGGAATATATGAGGTC	703

REFERENCE AUTHORS	TITLE	JOURNAL REFERENCE AUTHORS	TITLE	JOURNAL COMMENT	FEATURES	source
1 (bases 1 to 2462)	Sahin,U., Tuerceli,O., Eberle,T., Vollmar,E., Villena-Heinsen,C., Seitz,G. and Pfeundschnuh,M. A novel tumor-specific leucine zipper protein sharing features with the N-myc oncogene induces immune responses in tumor-bearing patients	2 (bases 1 to 2462)	Sahin,U., Tuerceli,O., Eberle,T., Vollmar,E., Villena-Heinsen,C., Seitz,G. and Pfeundschnuh,M. Direct Submission Submitted (21-JAN-1999) Internal Medicine, University of Saarland, Kirberg Str., Homburg, Saar 66421, Germany NCBI staff are still waiting for submitters to provide appropriate CDS information.		Location/Qualifiers 1. .2462 /organism="Homo sapiens" /db_xref="taxon:9606" /sex="male" /tissue_type="testis" 815...1951 /note="similar to HOM-TES-95 tumor antigen"	792 a 513 c 551 q 606 t
					misc_feature	
					BASE COUNT	
					GCING	

Query Match	82.9%;	Score 2140;	DB 9;	Length 2462;
Best Local Similarity	95.4%;	Pred. No. 0;		
Matches 2303;	Conservative	0;	Mismatches 45;	Indels 66;
Gaps	7;			
QY	228	AATTAGTAATCCCTACAGACAAAGTTTGTACAACAATTCGTTGACGGCACCAGCAAA	287	
Db	16	AGTAAGGAGAGGTTCTCATCAGAGGTTTGTACAACAATTCGTTGACGGCACCAGCAAA	75	
QY	288	CATATTTAGTGCAAAAGTCTTACCTTGTACAAATGATAACAGATTTGATCCTGCTGG	347	
Db	76	CATATTTAGTGCAAAAGTCTTACCTTGTACAAATGATAACAGATTTGATCCTGCTGG	135	
QY	348	AGATGAGTAATATTTTATACCAACGTTGACGAAGATGCAGAAACCAACAGACAATGCCA	407	
Db	136	AGATGGAGTAATATTTATACCAAGTTGACGAAGATGCAGAAACCAACAGACAATGCCA	195	
QY	408	ATTTACGTTGTCATATGGAACACTATTAGATATGACGTACCCAAATGACCTTACAC	467	
Db	196	ATTTACGTTGTCATATGGAACACTATTAGATATGACGTACCCAAATGACCTTACAC	255	
QY	468	TTTTCTCTCTTGGTGAAGA-TGGAACTGTTAGTGGTTTGATACACGCATCAAAACTA	526	
Db	256	TTTTCTCTCTTGGTGAAGA-TGGAACTGTTAGTGGTTTGATACACGCATCAAAACTA	315	
QY	527	GCTGCACAAAGAAGATTGAAAGATCAATTTAATTAACGTGCGAGTGTGCCAGT	586	
Db	316	GCTGCACAAAGAAGATTGAAAGATCAATTTAATTAACGTGCGAGTGTGCCAGT	375	
QY	587	CTGPT--GCPATTTGCCACCACCAATACCATATTA-CTTGCTGTTGGTTGTTCTGACAGCT	643	
Db	376	CTGTTTGTCTATTTTGGCCACCACATACCATTTACCTTGTGTTGGTTGTTCTGACAGCT	435	
QY	644	CAGTACGAATATATGATCGCGAATGCTGGGCACAAAGAGCTACAGGGAATTTATCAGGTC	703	
Db	436	CAGTACGAATATATGATCGCGAATGCTGGGCACAAAGAGCTACAGGGAATTTATCAGGTC	495	
QY	704	GAGGGACTACTGGAATGGTTGCCCTTTTATTCCTCCCATCTTTAATAAGTCTTGCA	763	
Db	496	GAGGGACTACTGGAATGGTTG-CGGTTTATTCCTCCCATCTTTAATAAGTCTTGCA	554	
QY	764	GAGTGACATCTCTGTGTACAGTGAAGATGATCAAGAGATTCCTGTAGTACTCTTCAG	823	
Db	555	GAGTGACATCTCTGTGTACAGTGAAGATGATCAAGAGATTCCTGTAGTACTCTTCAG	614	
QY	824	ATTACATATATCTTTTGACCCGAAAGATGATACAGCAGAGAACTTAAATCTCTTG	883	
Db	615	ATTACATATATCTTTTGACCCGAAAGATGATACAGCAGAGAACTTAAATCTCTTG	674	

[illegible]

QY	1907	CTGGGCGTTTCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGAGACACAGATG	1966
Db	1755	CTGGGCGTTTCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGAGACAGATG	1813
QY	1967	ACAGTGTGATGATACCCAGTCTTGATCCCAAGGTGCAAGGTATCGAGCAGGACCTGGTGATA	2026
Db	1814	ACAGTGTGATGATACCCAGTCTTGATCCCAAGGTGCAAGGTATCGAGCAGGACCTGGTGATA	1870
QY	2027	GAGCGTCTGCTGTGTGGCCGTTATTCAGGAGTCTTTCAGACGGAGGAAAAAGAAAGAAA	2086
Db	1871	GAGCGTCTGCTGTGTGGCCGTTATTCAGGAGTCTTTCAGACGGAGGAAAAAGAAAGAAA	1930
QY	2087	TGGAAGAAATTTGGATACATTTGAACATTTAGAAGCGCGCTAGTAAATAGTTTATTAAGGCC	2146
Db	1931	TGGAAGAAATTTGGATACATTTGAACATTTAGAAGCGCGCTAGTAAATAGTTTATTAAGGCC	1990
QY	2147	ATCGGCACTTCAGACACATGATAAAGAAAGCCAAATTTCTGGGTGCTAACTTTGTAATGA	2206
Db	1991	ATCGGCACTTCAGACACATGATAAAGAAAGCCAAATTTCTGGGTGCTAACTTTGTAATGA	2055
QY	2207	GTGGTCTTCGACTGTGGCCACATTTTCAATCTGGGATCGGCACACTGCTGAGCATTTGATGC	2266
Db	2051	GTGGTCTTCGACTGTGGCCACATTTTCAATCTGGGATCGGCACACTGCTGAGCATTTGATGC	2111
QY	2267	TTCTGGAGCTCATATCATGTGTGTAACCTGCCCTGCAGCCACACATCCGTTTGACCCAAATT	2321
Db	2111	TTCTGGAGCTCATATCATGTGTGTAACCTGCCCTGCAGCCACACATCCGTTTGACCCAAATT	217
QY	2327	TAGCCTTCATCTGCGATAGATTAATGACATTAAGATCTGCTCAACCATTAAGAAGAGTCAAGGA	238
Db	2171	TAGCCTTCATCTGCGATAGATTAATGACATTAAGATCTGCTCAACCATTAAGAAGAGTCAAGGA	2233
QY	2387	TTTTTTAACGGAAACCTTGCTGATGAAGTTTATAACTCGAAACGAACACTCATGCTGGAAGAAA	2444
Db	2231	TTTTTTAACGGAAACCTTGCTGATGAAGTTTATAACTCGAAACGAACACTCATGCTGGAAGAAA	2299
QY	2447	CTAGAAAACACATTCAGTTCAGGCTCTTTCATGTTGAGGATGTGGCTTCACCTTAATC	2505
Db	2291	CTAGAAAACACATTCAGTTCAGGCTCTTTCATGTTGAGGATGTGGCTTCACCTTAATC	2333
QY	2507	ATATCCGAGTTCACCGGTTTGAGGGGTGACAGATTCAGAAGGCTCTGGTCAAGAGATGAAA	2511
Db	2351	ATATCCGAGTTCACCGGTTTGAGGGGTGACAGATTCAGAAGGCTCTGGTCAAGAGATGAAA	2411
QY	2567	ATGAGGATGAGGAA	2580
Db	2411	ATGAGGATGAGGAA	2424

TITLE
Shibamura, I., Tanaka, T.
Direct Submission
Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
JOURNAL

```
Db 1081 CGAGCAACAGGATTAACAAATGAAAGCTGAGCCCAACAGGACAGGTGAACAGT 1140
QY 1428 TTAACTTTGCACTACAGACACAGAGCAACTACAGACACATATAAACTGAACCTTTAC 1487
Db 1141 TTAACTTTGCACTACAGACACAGAGCAACTACAGACACATATAAACTGAACCTTTAC 1200
QY 1488 AGATGAATGAGAGGATATAGCATCAAGTTCTAGAGGAATTTGGAGCAATTCGAATCTGA 1547
Db 1201 AGATGAATGAGAGGATATAGCATCAAGTTCTAGAGGAATTTGGAGCAATTCGAATCTGA 1260
QY 1548 GGGTCAGGAGGAATCTTTCCTCCACAGAGCTCACTGCAACCAACAGAGGACAGTGA 1607
Db 1261 GGGTCAGGAGGAATCTTTCCTCCACAGAGCTCACTGCAACCAACAGAGGACAGTGA 1320
QY 1608 AACAAAGCTCCTGAAGATCATCAGAGGATGTGCAAAATATCAGGAGGATCTGCG 1667
Db 1321 AACAAAGCTCCTGAAGATCATCAGAGGATGTGCAAAATATCAGGAGGATCTGCG 1380
QY 1668 AGAAACCCAGTTGAGAACCATATCAATATTAACACATTCAGATAAGTTTCACAGCAAGCC 1727
Db 1381 AGAAACCCAGTTGAGAACCATATCAATATTAACACATTCAGATAAGTTTCACAGCAAGCC 1440
QY 1728 ATTGATTTCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTGTGGGTTCC 1787
Db 1441 ATTGATTTCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTGTGGGTTCC 1500
QY 1788 AGAAGATCTGCTTCACTCTCAAAAGCCCAAGAACCAAGGAACTTCAGATCAGATGAC 1847
Db 1501 AGAAGATCTGCTTCACTCTCAAAAGCCCAAGAACCAAGGAACTTCAGATCAGATGAC 1560
QY 1848 TGAGAGTGCTTACCAATGAAATTAACCAATCTGAGCCCTCACTTCCAAACAGAGCCAC 1907
Db 1561 TGAGAGTGCTTACCAATGAAATTAACCAATCTGAGCCCTCACTTCCAAACAGAGCCAC 1620
QY 1908 TGGGCTCTCAGCTCATGAACAACTCCACAGGAGTCTGCTCTTTCAGGACACAGATGA 1967
Db 1621 TGGGCTCTCAGCTCATGAACAACTCCACAGGAGTCTGCTCTTTCAGGACACAGATGA 1680
QY 1968 CAGTGAATGATGACCCAGTCTGATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGAATG 2027
Db 1681 CAGTGAATGATGACCCAGTCTGATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGAATG 1740
QY 2028 AGCCTCTGCTGTGCTGCTTTCAGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2087
Db 1741 AGCCTCTGCTGTGCTGCTTTCAGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
QY 2088 GGAAGATTTGATATCTTTGAACATTTAGAGGCGCTAGTAAATGCTTTTAAAGGCCA 2147
Db 1801 GGAAGATTTGATATCTTTGAACATTTAGAGGCGCTAGTAAATGCTTTTAAAGGCCA 1860
QY 2148 TCGCAACTCCAGGACATGATTAAGAGCCAAATTTCTGGGGTGTCTACTTTGTAATGAG 2207
Db 1861 TCGCAACTCCAGGACATGATTAAGAGCCAAATTTCTGGGGTGTCTACTTTGTAATGAG 1920
QY 2208 TGGTTCGACTGTGGCCACATTTTCATCTGGGATCGGCACACTCTGAGCATTTGATGCT 2267
Db 1921 TGGTTCGACTGTGGCCACATTTTCATCTGGGATCGGCACACTCTGAGCATTTGATGCT 1980
QY 2268 TCTGGAAGCTGATTAATCATGTGTGTAATCTGCCTGACGCCACATCCGTTTGACCCAAATTT 2327
Db 1981 TCTGGAAGCTGATTAATCATGTGTGTAATCTGCCTGACGCCACATCCGTTTGACCCAAATTT 2040
QY 2328 AGCCTCATCTGGCATAGATTAATGATTAAGAGTCTGTCACCAATTAAGAGTCAAGGAT 2387
Db 2041 AGCCTCATCTGGCATAGATTAATGATTAAGAGTCTGTCACCAATTAAGAGTCAAGGAT 2100
QY 2388 TTTTAAACCGAAACTTGTGATGAAGTTTAACTCGAAAGCAACTCATGCTGCGAAGAAC 2447
Db 2101 TTTTAAACCGAAACTTGTGATGAAGTTTAACTCGAAAGCAACTCATGCTGCGAAGAAC 2160
QY 2448 TAGAAACACATTTACAGTTCCAGCCTCTTTCATGTTGAGGATTTGGCTTCACTTAATCA 2507
```

```
Db 2161 TAGAAACACACATTACAGTTCACGCTTCCAGCTTTTCATGTTGAGGATGTTGGCTTCACTTAATCA 2220
QY 2508 TATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAGGCTCTGGTCAAGAGATGAAGA 2567
Db 2221 TATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAGGCTCTGGTCAAGAGATGAAGA 2280
QY 2568 TGAGGATGAGGAA 2580
Db 2281 TGAGGATGAGGAA 2293
```

RESULT 10
BC025262

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC025262 2024 bp mRNA linear PRI 08-MAR-2002
IMAGE:4763336, mRNA, complete cds.

BC025262
MGC.
GI:19263812

Homo sapiens.
Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2024)
Strausberg, R.
Direct Submission

Submitted (05-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov

Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gulin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalrus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL plate: 42 Row: 9 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8923955.

Location/Qualifiers
1. .2024

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:39152 IMAGE:4763336"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH_MGC_48"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
58. .1599
/codon_start=1
/product="Similar to PC326 protein"
/protein_id="AAH25262.1"
/db_xref="GI:19263812"

translation="MLSRFEASEVAQSNRGRSRPRGGTQSQDSTLPTVPSPSPD
LEVSATMEVDTPAQFQPSFTSMTSAQAHSTSPSPSPSTPLLSPPDSRQSRQSV
ASGHHTHQSDNNNEKLSKPKGTGPVLSHYSTGTTTSTIKLFTDWSIASSSR
GTGSHCKSEGESEFVQSSVQPPGDSKAPESSESDATKIQGVSAAENVEHNH
ITQSKDKTAPLSDNSGRNDLNLDCGVPESSAKAKPEPSTDTSTESATNEN
NTNPPQTEATGTPSAHEETSRDLSQDTDDDDPVLIPGARYRAGPGRRSAVA

FEATURES
SOURCE

CDS

Db	901	CCTCAGTTC	CAACAGAAC	CCACTGG	CCCTT	CAGCTC	ATGATG	AGAAAC	ATCCAC	AGGAC	950
Qy	1945	TCTCCTCT	CAGGACAC	AGATGAC	AGTATG	ATGATG	ATGATG	ATGATG	ATGATG	ATGATG	2004
Db	961	TCTCCTCT	CAGGACAC	AGATGAC	AGTATG	ATGATG	ATGATG	ATGATG	ATGATG	ATGATG	1020
Qy	2005	TATGAGC	AGGACCT	GGTGTG	ATAGAC	GGCTCT	GTCTGT	TTCCCG	GTATTC	AGGAGT	2064
Db	1021	TATGAGC	AGGACCT	GGTGTG	ATAGAC	GGCTCT	GTCTGT	TTCCCG	GTATTC	AGGAGT	1080
Qy	2065	CGGAGAA	AGAAAG	AAAGAA	AAATG	GAATG	GAATG	GAATG	GAATG	GAATG	2124
Db	1081	CGGAGAA	AGAAAG	AAAGAA	AAATG	GAATG	GAATG	GAATG	GAATG	GAATG	1140
Qy	2125	GTAAATG	TTTATA	AAAGG	CCATC	CGCAAC	TCCAG	GACAAT	GATAAA	AGAGC	2184
Db	1141	GTAAATG	TTTATA	AAAGG	CCATC	CGCAAC	TCCAG	GACAAT	GATAAA	AGAGC	1200
Qy	2185	TGGGGT	GTCTAA	CTTTGT	TAATG	AGTGG	TCTG	ACTG	TGCG	CACTTT	2244
Db	1201	TGGGGT	GTCTAA	CTTTGT	TAATG	AGTGG	TCTG	ACTG	TGCG	CACTTT	1260
Qy	2245	CACACTG	CTGAG	CAATTT	GATGCT	TTGGA	AGCTG	ATATG	ATATG	ATATG	2304
Db	1261	CACACTG	CTGAG	CAATTT	GATGCT	TTGGA	AGCTG	ATATG	ATATG	ATATG	1320
Qy	2305	CCACATC	CGGTTG	ACCCAA	TTTATG	CTCTG	GAAGC	TGATG	ATATG	ATATG	2364
Db	1321	CCACATC	CGGTTG	ACCCAA	TTTATG	CTCTG	GAAGC	TGATG	ATATG	ATATG	1380
Qy	2365	TCACCAT	TGAAG	AGTCA	AGGAT	TTTAA	CCGAA	AACTT	GTCTG	ATATG	2424
Db	1381	TCACCAT	TGAAG	AGTCA	AGGAT	TTTAA	CCGAA	AACTT	GTCTG	ATATG	1440
Qy	2425	AACGAAC	TATGCT	GGAGAA	AACTG	ATATG	ATATG	ATATG	ATATG	ATATG	2484
Db	1441	AACGAAC	TATGCT	GGAGAA	AACTG	ATATG	ATATG	ATATG	ATATG	ATATG	1500
Qy	2485	AGGATG	TTGGCT	TTCAC	TTAAT	CAATC	CGAGC	TGAC	CGGTG	GGAGT	2544
Db	1501	AGGATG	TTGGCT	TTCAC	TTAAT	CAATC	CGAGC	TGAC	CGGTG	GGAGT	1560
Qy	2545	GGCTCT	GGTCA	AGAGAT	GAAAT	GAGAT	GAGAA	GAGAA	GAGAA	GAGAA	2580
Db	1561	GGCTCT	GGTCA	AGAGAT	GAAAT	GAGAT	GAGAA	GAGAA	GAGAA	GAGAA	1596
RESULT 11	HSR01706	1952 bp	mRNA	linear	PRI 20-MAR-2002						
LOCUS	HSR01706	1952 bp	mRNA	linear	PRI 20-MAR-2002						
DEFINITION	complete cds.										
ACCESSION	AL136738										
VERSION	AL136738.1										
KEYWORDS	GI:12052994										
SOURCE	human.										
ORGANISM	Homo sapiens										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.										
AUTHORS	Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.										
TITLE	Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152										
JOURNAL	Martinsried, GERMANY										
COMMENT	Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZ434C211) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available										

BASE COUNT	670 a	392 c	438 g	524 t							
ORIGIN											
Query Match	61.8%;	Score 1594.4;	DB 9;	Length 2024;							
Best Local Similarity	99.9%;	Pred. No. 0;									
Matches 1595;	Conservative	0; Mismatches	1; Indels	0; Gaps	0;						
Qy	985	CGAGACG	AGATG	AGAGC	AGATG	CCCAAT	GTGTCA	TATGAT	GCAGAG	AAATG	1044
Db	1	CGAGACG	AGATG	AGAGC	AGATG	CCCAAT	GTGTCA	TATGAT	GCAGAG	AAATG	60
Qy	1045	TTATCA	AGATG	TTGA	AGACG	AGATG	AGATG	AGATG	AGATG	AGATG	1104
Db	61	TTATCA	AGATG	TTGA	AGACG	AGATG	AGATG	AGATG	AGATG	AGATG	120
Qy	1105	TCTGAC	CCAGAG	TGGAA	CAATCA	TATAT	TTCA	ACTCT	CTTCC	TACGGT	1164
Db	121	TCTGAC	CCAGAG	TGGAA	CAATCA	TATAT	TTCA	ACTCT	CTTCC	TACGGT	180
Qy	1165	AGTCTG	ATTTG	GAAGT	GAATG	CAATG	CAATG	CAATG	CAATG	CAATG	1224
Db	181	AGTCTG	ATTTG	GAAGT	GAATG	CAATG	CAATG	CAATG	CAATG	CAATG	240
Qy	1225	CTTACG	CTTCT	TACAT	CTCT	TACAT	CTCT	TACAT	CTCT	TACAT	1284
Db	241	CTTACG	CTTCT	TACAT	CTCT	TACAT	CTCT	TACAT	CTCT	TACAT	300
Qy	1285	GAAAGC	CCCTCA	TTCTG	CTCT	TTCTG	CTCT	TTCTG	CTCT	TTCTG	1344
Db	301	GAAAGC	CCCTCA	TTCTG	CTCT	TTCTG	CTCT	TTCTG	CTCT	TTCTG	360
Qy	1345	GAGGATC	TGGAC	ACACAC	ACATC	ATGCT	CTGAT	CAACA	TATAT	GAAAGT	1404
Db	361	GAGGATC	TGGAC	ACACAC	ACATC	ATGCT	CTGAT	CAACA	TATAT	GAAAGT	420
Qy	1405	AAACAG	GACAG	GGTGA	ACAGG	TTTAA	CTTGC	ACTAC	GACAG	AGAGG	1464
Db	421	AAACAG	GACAG	GGTGA	ACAGG	TTTAA	CTTGC	ACTAC	GACAG	AGAGG	480
Qy	1465	AGCACAT	ATAACT	GAAC	TTTAC	AGATG	ATGAT	GAGAG	CAATG	ATGAT	1524
Db	481	AGCACAT	ATAACT	GAAC	TTTAC	AGATG	ATGAT	GAGAG	CAATG	ATGAT	540
Qy	1525	ATTGGG	AGCCAT	TGCA	ATCTG	AGGCTC	AGGAG	GAATC	TTTCT	CCACAG	1584
Db	541	ATTGGG	AGCCAT	TGCA	ATCTG	AGGCTC	AGGAG	GAATC	TTTCT	CCACAG	600
Qy	1585	CAACCA	CCAGAG	GACAG	TGAAC	AAAGCT	CCTGA	AGATC	ATCAG	AGGAT	1644
Db	601	CAACCA	CCAGAG	GACAG	TGAAC	AAAGCT	CCTGA	AGATC	ATCAG	AGGAT	660
Qy	1645	AAATAT	CAGGAG	AGATG	ATCTG	CAAGAA	CCAGT	TGAG	CAACCAT	ATATAT	1704
Db	661	AAATAT	CAGGAG	AGATG	ATCTG	CAAGAA	CCAGT	TGAG	CAACCAT	ATATAT	720
Qy	1705	TCAGAT	AGTTT	CACG	CAAGC	CAATG	GGAT	TCCAA	CTCAG	GAGAA	1764
Db	721	TCAGAT	AGTTT	CACG	CAAGC	CAATG	GGAT	TCCAA	CTCAG	GAGAA	780
Qy	1765	CTTGAT	CGCTCT	TTGG	GGTTCC	AGAGAA	TTCTG	CTTCA	CTCTG	AAAGCC	1824
Db	781	CTTGAT	CGCTCT	TTGG	GGTTCC	AGAGAA	TTCTG	CTTCA	CTCTG	AAAGCC	840
Qy	1825	GAACTT	CAGAT	CAGAT	AGTCT	AGTCT	AGTCT	AGTCT	AGTCT	AGTCT	1884
Db	841	GAACTT	CAGAT	CAGAT	AGTCT	AGTCT	AGTCT	AGTCT	AGTCT	AGTCT	900
Qy	1885	CCTCAG	TTCCAA	ACAG	AGCC	ACTGG	CGCTT	CAGT	CTATG	AGAAC	1944

QY	1483	TTTACAGATGAATGGAGCAGTATAGCATCAAGTCTAGAGGAATGGAGCCATTGGCAAA	1542
Db	481	TTTACAGATGAATGGAGCAGTATAGCATCAAGTCTAGAGGAATGGAGCCATTGGCAAA	540
QY	1543	TCTGAGGCTCAGAGGAATCTTTGCTCCACAGAGCTCAGTGCACACCACAGAGGAGAC	1602
Db	541	TCTGAGGCTCAGAGGAATCTTTGCTCCACAGAGCTCAGTGCACACCACAGAGGAGAC	600
QY	1603	AGTGAACAAAAGCTCCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAGGAGTA	1662
Db	601	AGTGAACAAAAGCTCCTGAAGATCATCAGAGGATGTGACAAAATATCAGGAGGAGTA	660
QY	1663	TCTGCAGAAACCCAGTTGAGAACCATATCAATATAAACAATCAGATAGTTCACAGCC	1722
Db	661	TCTGCAGAAACCCAGTTGAGAACCATATCAATATAAACAATCAGATAGTTCACAGCC	720
QY	1723	AAGCCATTGGATTCCAACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGCGG	1782
Db	721	AAGCCATTGGATTCCAACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGCGG	780
QY	1783	GTTCAGAGAAATCTGCTTCATCTGAAAAGCCAAAGGAACCAAGAACTTCAGATCAGACT	1842
Db	781	GTTCAGAGAAATCTGCTTCATCTGAAAAGCCAAAGGAACCAAGAACTTCAGATCAGACT	840
QY	1843	AGCACTGAGAGTGTACCAATGAAAATAACACCAATCTGAGCCCTAGTTCCAAACAGAA	1902
Db	841	AGCACTGAGAGTGTACCAATGAAAATAACACCAATCTGAGCCCTAGTTCCAAACAGAA	900
QY	1903	GCCACTGGGCTTCAGCTCATGAAGAAACATCCACAGGACTCTGCTCTTCAGGACACA	1962
Db	901	GCCACTGGGCTTCAGCTCATGAAGAAACATCCACAGGACTCTGCTCTTCAGGACACA	960
QY	1963	GATGACAGTGATGATGACCCAGTCTGATCCAGGTGCAAGGTATCAGACGAGCCTGGT	2022
Db	961	GATGACAGTGATGATGACCCAGTCTGATCCAGGTGCAAGGTATCAGACGAGCCTGGT	1020
QY	2023	GATGACGCTCTGCTGTGGCCGTATTCAGAGTCTTCAGAGCGAGAAAGAAAGGAAA	2082
Db	1021	GATGACGCTCTGCTGTGGCCGTATTCAGAGTCTTCAGAGCGAGAAAGAAAGGAAA	1080
QY	2083	GAAATGGAAGAAATGGATACTTTGAACATTAGAAGGCCGCTAGTAAAAATGGTTATAAA	2142
Db	1081	GAAATGGAAGAAATGGATACTTTGAACATTAGAAGGCCGCTAGTAAAAATGGTTATAAA	1140
QY	2143	GGCCATCGCACTCCAGGACAATGATAAAGAAAGCAATTTCTGGGTGCTAACTTTGTA	2202
Db	1141	GGCCATCGCACTCCAGGACAATGATAAAGAAAGCAATTTCTGGGTGCTAACTTTGTA	1200
QY	2203	ATGAGTGTCTGACTGTGGCCACATTTTCATCTGGATCGGCACACTGCTGAGCAATTG	2262
Db	1201	ATGAGTGTCTGACTGTGGCCACATTTTCATCTGGATCGGCACACTGCTGAGCAATTG	1260
QY	2263	ATGCTTCTGGAAGCTGATAATCATGTGGTAAACTGCCTCAGCCACATCCGTTTGACCCA	2322
Db	1261	ATGCTTCTGGAAGCTGATAATCATGTGGTAAACTGCCTCAGCCACATCCGTTTGACCCA	1320
QY	2323	ATTTTAGCCCTCATCTGGCATAGATTATGACATAAAGATCTGCTCACCATTAGAAGAGTCA	2382
Db	1321	ATTTTAGCCCTCATCTGGCATAGATTATGACATAAAGATCTGCTCACCATTAGAAGAGTCA	1380
QY	2383	AGGATTTTAAACCAAACTTGCTGATGAAGTTATACTCGAAAGCAACTCATGCTGGAA	2442
Db	1381	AGGATTTTAAACCAAACTTGCTGATGAAGTTATACTCGAAAGCAACTCATGCTGGAA	1440
QY	2443	GAAACTAGAAACCACTTACAGTTCAGCCCTTTTCATGTTTCAGATGTTGGCTTCACATT	2502
Db	1441	GAAACTAGAAACCACTTACAGTTCAGCCCTTTTCATGTTTCAGATGTTGGCTTCACATT	1500
QY	2503	AATCATATCCGAGCTGACCGGTTGGAGGTGACAGATCAGAAGGCTCTGGTCAAGAGAT	2562
Db	1501	AATCATATCCGAGCTGACCGGTTGGAGGTGACAGATCAGAAGGCTCTGGTCAAGAGAT	1560


```

|||||
Db 1561 GAAATGAGGATGAGGAA 1578

RESULT 12
AK093970
LOCUS
DEFINITION
Homo sapiens cDNA FLJ36651 fis, clone UTERU2000095, highly similar
to Homo sapiens PC326 protein (PC326) mRNA.
ACCESSION
AK093970
VERSION
AK093970.1 GI:21752937
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens uterus cDNA to mRNA, clone lib:UTERU2
clone:UTERU2000095.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS
Ninomiya,K., Nagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Sugiyama,T., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Yamamoto,J., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project

TITLE
Unpublished
JOURNAL
2 (bases 1 to 2803)
REFERENCE
Isogai,T. and Yamamoto,J.
AUTHORS
Direct Submission
JOURNAL
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
source
1..2803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UTERU2000095"
/tissue_type="uterus"
/clone_lib="UTERU2"
/note="cloning vector: pME18SFL3"
BASE COUNT 785 a 608 c 639 g 771 t
ORIGIN
Query Match 53.4%; Score 1378.4; DB 9; Length 2803;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATGCTCGGGTGGCTCTACCCACACCTGTGTGGGACGTGAGGAAAAGGTCCTCGGG 60
Db 240 ATGCTCGGGTGGCTCTACCCACACCTGTGTGGGACGTGAGGAAAAGGTCCTCGGG 299
QY 61 CTGGAGGACCGTCCCGGCTGCGGAGTCGCTACCTGGGAGAGAGAGATTTATCCAAAGA 120
Db 300 CTGGAGGACCGTCCCGGCTGCGGAGTCGCTACCTGGGAGAGAGAGATTTATCCAAAGA 359
QY 121 TAAACTTTGAACACCCCTTAATGTGCATGATGTTGTGTATATACAAATCTGTGGAAT 180
Db 360 TAAAACTTTGAACACCCCTTAATGTGCATGATGTTGTGTATATACAAATCTGTGGAAT 419
QY 181 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCCT 240
Db 420 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCCT 479

```

```

QY 241 TACAGCAAGAGGTTTGGACAACAATTCGTTTCAGGGCACCGAGCAACAACATATTACTGCA 300
Db 480 TACAGCAAGAGGTTTGGACAACAATTCGTTTCAGGGCACCGAGCAACAACATATTACTGCA 539
QY 301 AAGTCTTACCTTGTACAAATGATATAACAGATTCCTGCTGCTGGAGATGGAGTAATA 360
Db 540 AAGTCTTACCTTGTACAAATGATATAACAGATTCCTGCTGCTGGAGATGGAGTAATA 599
QY 361 TTTTATACCAACGTTGAGCAGATGCGAAGAACCAACAGCAATGCCAATTTACGTGTCAT 420
Db 600 TTTTATACCAACGTTGAGCAGATGCGAAGAACCAACAGCAATGCCAATTTACGTGTCAT 659
QY 421 TATGGAACCTACTTATGAGATTATGACTGTACCCCAATGACCCCTTACCTTTCTCTTGT 719
Db 660 TATGGAACCTACTTATGAGATTATGACTGTACCCCAATGACCCCTTACCTTTCTCTTGT 719
QY 481 GGTGAAGATGGAACCTGTTAGTGGTTTCATACGCAATCAAAACTAGTCGCACAAAAGAA 540
Db 720 GGTGAAGATGGAACCTGTTAGTGGTTTCATACGCAATCAAAACTAGTCGCACAAAAGAA 779
QY 541 GATTGTAAAGATGATATTTTAACTGTCGACGTGTCGACGCTGCTGTTGCTATTTGC 600
Db 780 GATTGTAAAGATGATATTTTAACTGTCGACGTGTCGACGCTGCTGTTGCTATTTGC 839
QY 601 CCACCAATACCAATATTTACCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
Db 840 CCACCAATACCAATATTTACCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 899
QY 661 CGCGGAATGCTGGGCAACAAGAGCTACAGGCAATATATGAGGTCGAGGGGACTACTGGAATG 720
Db 900 CGCGGAATGCTGGGCAACAAGAGCTACAGGCAATATATGAGGTCGAGGGGACTACTGGAATG 959
QY 721 GTTGCCCGCTTTTATTCCTTCCCATCTTATATTAAGTCTGCGAGAGTGACATCTCTGCT 780
Db 960 GTTGCCCGCTTTTATTCCTTCCCATCTTATATTAAGTCTGCGAGAGTGACATCTCTGCT 1019
QY 781 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTACTCTTCAGATTACATATATCTTTT 840
Db 1020 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTACTCTTCAGATTACATATATCTTTT 1079
QY 841 GACCCGAAAGATGATACAGCAGCAAGACTTAAACCTCTTCGGAAGAGAGAGAGAA 900
Db 1080 GACCCGAAAGATGATACAGCAGCAAGACTTAAACCTCTTCGGAAGAGAGAGAGAA 1139
QY 901 GAGTTCGCAACCAACCACTTAAGCGTTTGGAGACTTCGTTGATGGTTCAGATACCTGGA 960
Db 1140 GAGTTCGCAACCAACCACTTAAGCGTTTGGAGACTTCGTTGATGGTTCAGATACCTGGA 1199
QY 961 CCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGCGAGAGAGAGATGCCAATGTGCA 1020
Db 1200 CCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGCGAGAGAGAGATGCCAATGTGCA 1259
QY 1021 TTGATGAGAGAAATGTCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGAGTTGCA 1080
Db 1260 TTGATGAGAGAAATGTCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGAGTTGCA 1319
QY 1081 CAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAGTCAATCAGATATT 1140
Db 1320 CAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAGTCAATCAGATATT 1379
QY 1141 TCAACTCTTCTTACCGTCCCATCAAGTCTGATTTTGAAGTGAAGTGAAGTGAAGTGAAG 1200
Db 1380 TCAACTCTTCTTACCGTCCCATCAAGTCTGATTTTGAAGTGAAGTGAAGTGAAGTGAAG 1439
QY 1201 GTAGATATCTTCCAGCTGAACAATTTCTCAGCGCTTCTACATCTCTCAATGTCACTCAG 1260
Db 1440 GTAGATATCTTCCAGCTGAACAATTTCTCAGCGCTTCTACATCTCTCAATGTCACTCAG 1499
QY 1261 GCTCATCTCAGATCATCTCCACAGAAAGCCCTCATTTCTACTCTTCTGCTATCTTCTCCA 1320
Db 1500 GCTCATCTCAGATCATCTCCACAGAAAGCCCTCATTTCTACTCTTCTGCTATCTTCTCCA 1559

```



```
repeat_region /rpt_family="L1M4" 12902..13207
repeat_region /rpt_family="AluJo" 13213..13516
repeat_region /rpt_family="AluSx" 13600..13795
repeat_region /rpt_family="L1M4" 13796..14103
repeat_region /rpt_family="AluY" 14104..15110
repeat_region /rpt_family="L1M4" 15111..15412
repeat_region /rpt_family="AluSg" 15413..15877
repeat_region /rpt_family="L1M4" 15867..15963
repeat_region /rpt_family="L1M4" 15953..16049
repeat_region /rpt_family="L1M4" 16050..16766
repeat_region /rpt_family="L1MB8" 16823..17033
repeat_region /rpt_family="MER2" 17034..17080
repeat_region /rpt_family="(TTTTTA)n" complement(17324..17625)
repeat_region /rpt_family="AluSx" complement(18084..18138)
repeat_region /rpt_family="MIR" complement(18566..18706)
repeat_region /rpt_family="MIR" 19349..19415
repeat_region /rpt_family="(CAGAGA)n" complement(20636..21728)
repeat_region /rpt_family="L1MD1" complement(21729..22087)
repeat_region /rpt_family="MLT1A1" complement(22088..22269)
repeat_region /rpt_family="L1MD1" complement(22270..22429)
repeat_region /rpt_family="Alu3b" complement(22488..22840)
repeat_region /rpt_family="L1MD1" 23456..23764
repeat_region /rpt_family="AluY" complement(23767..24338)
repeat_region /rpt_family="L1MD1" 23842
unsure /note="Probably C, possibly A."
repeat_region complement(24635..24942)
repeat_region /rpt_family="AluSg" complement(25255..25293)
repeat_region /rpt_family="L2" 25294..25575
repeat_region /rpt_family="AluSx" complement(25576..25720)
repeat_region /rpt_family="L2" complement(26187..26471)
repeat_region /rpt_family="AluSx" complement(26861..27080)
repeat_region /rpt_family="AluSg/X" complement(27508..28893)
repeat_region /rpt_family="pTR5" 28894..29189
repeat_region /rpt_family="AluSx"

Query Match 49.9%; Score 1287.2; DB 9; Length 162314;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1322; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGTCCTCGGGGTGGCTCCTGCCACATCTGTTGGGACATGAGGAAAAGGTCCTCCGCGG 60
Db 103593 ATGTCCTCGGGGTGGCTCCTGCCACATCTGTTGGGACATGAGGAAAAGGTCCTCCGCGG 103534
```

QY 1141 TCAACTCTTCTAGGTCCTCCATCAAGTCTCTGATTGGAGTGAGTGAACCTGCAATGGAA 1200
 Db 102453 TCAACTCTTCTAGGTCCTCCATCAAGTCTCTGATTGGAGTGAGTGAACCTGCAATGGAA 102394
 QY 1201 GTAGATCTCCAGCTGACGATTTCTTCAGCCCTTACATCCCTCTACAAATGTGAGTCTCAG 1260
 Db 102393 GTAGATCTCCAGCTGACGATTTCTTCAGCCCTTACATCCCTCTACAAATGTGAGTCTCAG 102334
 QY 1261 GCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTCTTGTCTATCTTCTCCCA 1320
 Db 102333 GCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTCTTGTCTATCTTCTCCCA 102274
 QY 1321 GACAGTGAACAAGGAGCTCTGTGAGGATCTGGACATCTGGACACACATCATCATCTGAT 1380
 Db 102273 GACAGTGAACAAGGAGCTCTGTGAGGATCTGGACATCTGGACACACATCATCATCTGAT 102214

RESULT 14
 AC016484
 LOCUS
 DEFINITION Homo sapiens chromosome 15 clone RP11-18H24, WORKING DRAFT
 AC016484
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 159098)
 Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,
 Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
 Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,
 Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,
 Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.
 Unpublished
 2 (bases 1 to 159098)
 Bruno, D., Conn, L., Della Rosa, M., Federspiel, N., Foreman, P.,
 Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
 Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
 Yu, S. and Davis, R.W.
 Direct Submission
 Submitted (01-DEC-1999) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On Mar 14, 2001 this sequence version replaced gi:12331479.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSLDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: 720
 Center clone name: RP11-18H24
 ----- Summary Statistics
 Sequencing Vector: M13mp18; X02513; 99% of reads
 Sequencing Vector: Plasmid; plasmid_accession; 1% of reads
 Chemistry: Dye-terminator; 1% of reads
 Assembly: Dye-terminator Big Dye; 99% of reads
 Consensus quality: 157616 bases at least Q40
 Consensus quality: 158170 bases at least Q30
 Consensus quality: 158357 bases at least Q20
 Insert size: 165454; agarose-fp
 Insert size: 158898; sum-of-contigs
 Quality coverage: 7.9x in Q20 bases; agarose-fp
 Quality coverage: 8.2x in Q20 bases; sum-of-contigs.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2166: contig of 2166 bp in length
 * 2167: gap of unknown length
 * 2267: 44132: contig of 41866 bp in length
 * 44133: 44232: gap of unknown length
 * 44233 159098: contig of 114866 bp in length.

FEATURES

source
 1. .159098
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /clone="RP11-18H24"
 /clone_lib="RPC1 human BAC library 11"
 1. 2166
 /note="assembly_name:Contig31"
 2267. .44132
 /note="assembly_name:Contig32"
 44233. .159098
 /note="assembly_name:Contig33"
 clone_end:SP6"
 BASE COUNT 46075 a 33092 c 33649 g 46082 t 200 others
 ORIGIN

Query Match 49.8%; Score 1285.6; DB 2; Length 159098;
 Best Local Similarity 95.7%; Pred. NO. 0;
 Matches 1321; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 QY 1 ATGCTCTGGGTGGCTCTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 60
 Db 109281 ATGCTCTGGGTGGCTCTCGCCACATCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 109340
 QY 61 CTGGAGGACCGTCCCGGCTCGGAGTCCCTACCTGGGAAGAGAGATTTATCAAGA 120
 Db 109341 CTGGAGGACCGTCTCTGGCTCTGGGCGCACCTACCTGGGAAGAGAGATTTATCAAGA 109400
 QY 121 TTAACACTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTTAATACATCTGTGGAAT 180
 Db 109401 TTGAACCTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTTAATACATCTGTGGAAT 109460
 QY 181 GACACTGAGAATATATTTATCTGGCTCAGATGACACCAATAGTAATAGTAATCT 240
 Db 109461 GACACTGAGAATATATTTATCTGGCTCAGATGACACCAATAGTAATAGTAATCT 109520
 QY 241 TACAGCAGAAAGTTTGGACAACATTCGTTGAGGCGCCAGCAACATATTTAGTGCA 300
 Db 109521 TACAGCAGAAAGTTTGGACAACATTCGTTGAGGCGCCAGCAACATATTTAGTGCA 109580
 QY 301 AAGTTCTTACCTTGTACAAATGATAACAGATGTATCTCTGCTCTGGAGATGAGTAATA 360
 Db 109581 AAGTTCTTACCTTGTACAAATGATAACAGATGTATCTCTGCTCTGGAGATGAGTAATA 109640
 QY 361 TTTTATACCAAGCTTGGAGAGATGAGCAACCAACAGCAATGCCAATTTAGCTGTCAT 420
 Db 109641 TTTTATACCAAGCTTGGAGAGATGAGCAACCAACAGCAATGCCAATTTAGCTGTCAT 109700
 QY 421 TATGGAATCTACTTATGAGATTTAGCTGTACCAATGACCCCTTACACTTTCTCTTGT 480
 Db 109701 AATGGAATCTACTTATGAGATTTAGCTGTACCAATGACCCCTTACACTTTCTCTTGT 109760
 QY 481 GCTGAAGTGAACCTGTTAGTGGTTGATACCGCATCAAACTACCTGCAACAAAGAA 540
 Db 109761 GCTGAAGTGAACCTGTTAGTGGTTGATACCGCATCAAACTACCTGCAACAAAGAA 109820
 QY 541 GATTTAAAGATGATATTTTAAATTAACCTGTCACCTGCTGCCACGTCTGCTATTTGC 600
 Db 109821 GATTTAAAGATGATATTTTAAATTAACCTGTCACCTGCTGCCACGTCTGCTATTTGC 109880
 QY 601 CCACCAATACCATACTCTGCTGTTGGTTGTTCTGACAGCTCAGTACGATATATGAT 660
 Db 109881 CCACCAATACCATACTCTGCTGTTGGTTGTTCTGACGTTGTTCTTATAGCTCAGTAAATATGAT 109940

Db 900 CTTTGAACATTAGAAGCCGCTAGTAAATGTTTATAAAGCCATCCCACTCAGGA 959
 QY 2162 CAATGATAAAGAGCAATTTCTGGGTGCTAACTTTGTAATGAGTGGTTCFAGCTGTG 2221
 Db 960 CAATGATAAAGAGCAATTTCTGGGTGCTAACTTTGTAATGAGTGGTTCFAGCTGTG 1019
 QY 2222 GCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCGGAAGCTGATA 2281
 Db 1020 GCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCGGAAGCTGATA 1079
 QY 2282 ATCATGTGTGTAACCTGCTGAGCCACATCCGTTTGACCAATTTAGCCTCATCTGGCA 2341
 Db 1080 ATCATGTGTGTAACCTGCTGAGCCACATCCGTTTGACCAATTTAGCCTCATCTGGCA 1133
 QY 2342 TAGATTATGACATAAAGATCTGGTCACCATAGAGAGTCAAGGATTTTAAACCGAAGAC 2401
 Db 1134 TAGATTATGACATAAAGATCTGGTCACCATAGAGAGTCAAGGATTTTAAACCGAAGAC 1193
 QY 2402 TTGCTGATGAAGTTTATACTGGAACGAACTCATGCTGGAAGAACTAGAACACCATTA 2461
 Db 1194 TTGCTGATGAAGTTTATACTGGAACGAACTCATGCTGGAAGAACTAGAACACCATTA 1253
 QY 2462 CAGTTCAGCCTCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGACC 2521
 Db 1254 CAGTTCAGCCTCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGACC 1313
 QY 2522 GGTGGAGGTGACAGATCAGAGGCTCTGGTCAAGAGATGAAATGAGGATGAGGAA 2580
 Db 1314 GGTGGAGGTGACAGATCAGAGGCTCTGGTCAAGAGATGAAATGAGGATGAGGAA 1372

RESULT 16
 AC107981
 LOCUS
 DEFINITION Homo sapiens chromosome 15 clone CTD-3247F8 map 15, LOW-PASS
 ACCESSION AC107981
 VERSION
 KEYWORDS HTG; HTGS; PHASEO.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 65517)
 Birren,B., Linton,L., Nusbaum,C., Allen,A., Allen,N.,
 Brown,A., Cantata,J., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Fierella,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,K., Jones,C.,
 Kamat,A., Karatas,A., Kellis,C., Larocque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
 Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
 Mienda,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tropham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Submitted
 Direct Submission
 TITLE
 JOURNAL
 REFERENCE

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L24536
 Center clone name: 3247_F_8

* NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 665: contig of 665 bp in length
 766 765: gap of 100 bp
 1461 1560: contig of 695 bp in length
 1561 2254: contig of 694 bp in length
 2255 2354: gap of 100 bp
 2355 3019: contig of 665 bp in length
 3020 3119: gap of 100 bp
 3120 3761: contig of 642 bp in length
 3762 3861: gap of 100 bp
 3862 4525: contig of 664 bp in length
 4526 4625: gap of 100 bp
 4626 5280: contig of 655 bp in length
 5281 5380: gap of 100 bp
 5381 6092: contig of 712 bp in length
 6093 6192: gap of 100 bp
 6193 6884: contig of 692 bp in length
 6885 6984: gap of 100 bp
 6985 7703: contig of 719 bp in length
 7704 7803: gap of 100 bp
 7804 8468: contig of 665 bp in length
 8469 8568: gap of 100 bp
 8569 9249: contig of 681 bp in length
 9250 9349: gap of 100 bp
 9350 10034: contig of 685 bp in length
 10035 10134: gap of 100 bp
 10135 10797: contig of 663 bp in length
 10798 10897: gap of 100 bp
 10898 11584: contig of 687 bp in length
 11585 11684: gap of 100 bp
 11685 12381: contig of 697 bp in length
 12382 12481: gap of 100 bp
 12482 13169: contig of 688 bp in length
 13170 13269: gap of 100 bp
 13270 13973: contig of 704 bp in length
 13974 14073: gap of 100 bp
 14074 14774: contig of 701 bp in length
 14775 14874: gap of 100 bp
 14875 15569: contig of 695 bp in length
 15570 15669: gap of 100 bp
 15670 16331: contig of 662 bp in length
 16332 16431: gap of 100 bp
 16432 17142: contig of 711 bp in length
 17143 17242: gap of 100 bp
 17243 17916: contig of 674 bp in length
 17917 18016: gap of 100 bp
 18017 18704: contig of 688 bp in length
 18705 18804: gap of 100 bp
 18805 19462: contig of 658 bp in length
 19463 19562: gap of 100 bp

DB 22371 GGTCAGGATATCGACGACGCCTGTGGTAG 22402

RESULT 18
AC026140
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
COMMENT

AC026140 163230 bp DNA linear HTG 26-MAY-2000
Homo sapiens chromosome 1 clone RP11-300C7 map 1, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC026140 GI:8077102
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163230)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 1, clone RP11-300C7
Unpublished
2 (bases 1 to 163230)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckghalter,B., Brown,A., Burkett,G.,
Campomiano,A., Castle,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Collimore,A., Cooke,P., Desellano,K., P., FitzHugh,W., Gage,D.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Goyette,M., Graham,L.,
Galagan,J., Gardyna,S., Glade,S., Glinder,S., Heaford,A., Horton,L.,
Grand-pierre,N., Grant,G., Hagos,B., Jones,C., Kann,L., Karatas,A.,
Howland,J.C., Iliev,I., Johnson,R., Landers,T., Lehoczyk,J.,
Klein,J., Larocque,K., Lamazares,R., McKernan,K., McPheeters,R.,
McCarthy,M., McEwan,P., McGurk,A., Miranda,C., Mlenga,V., Morrow,J.,
Melgrim,T., Menues,L., Mihova,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,S., Severy,P., Spencer,B., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,N., Subramanian,A., Talamas,J.,
Stange-Thomann,N., Stojanovic,N., Travers,M., Trigilio,J.,
Tsfaye,S., Theodore,J., Tirrell,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Vassiliev,H., Viel,R., Vo,A., Zimmer,A. and Zody,M.

Direct Submission
Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 23, 2000 this sequence version replaced gi:7523800.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8103
Center clone name: 300_C_7
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 152621 bases at least Q40
Consensus quality: 157889 bases at least Q30
Consensus quality: 159998 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 161630; sum-of-contigs
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

703H14 is from the library RPC14 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2>.

IMPORTANT: This sequence is not the entire insert of clone 703H14. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 702J19 is at 67710 in this sequence. The true right end of clone 295C6 is at 100 in this sequence.

FEATURES

source

```

1. .74729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q23.2-24.3"
/clone="RP4-703H14"
/clone_lib="RPC1-4"
/misc_feature
5. .112
/feature="L1M2 repeat: matches 1519. .1625 of consensus"
457. .754
/feature="AluX repeat: matches 2. .299 of consensus"
824. .935
/feature="L2 repeat: matches 2629. .2750 of consensus"
1050. .1570
/feature="L1M4 repeat: matches 3. .519 of consensus"
2795. .3025
/feature="L2 repeat: matches 2492. .2750 of consensus"
3038. .3096
/feature="MIR repeat: matches 192. .242 of consensus"
3097. .3399
/feature="AluX repeat: matches 1. .303 of consensus"
3400. .3553
/feature="MIR repeat: matches 3. .192 of consensus"
Join(3604. .3665,18433. .18525,26634. .26819,39299. .39412,
43008. .43143,45030. .45244,54286. .54379,55674. .55793,
56337. .56597,67585. .67763)
/feature="dJ703H14.1"
Join(<3604. .3665,18433. .18525,26634. .26819,39299. .39412,
43008. .43143,45030. .45244,54286. .54379,55674. .55793,
56337. .56597,67585. .67763)
/feature="dJ703H14.1"
/feature="match: ESTs: Em:N54388 Em:R14260 Em:H08089
Em:Z24943 Em:AAL30053"
/codon_start=3
/evidence="not_experimental"
/protein_id="G1:5725427"
/db_xref="GI:5725427"
/translating="REF1QRLKLEATLVNHDGCVNTICWDTGTYILSGSDTKLVI
SNPGRKVLITRSGRANIFAKELPCTNDKQIVSCSDGVIFVTNVEQADTNRC
QFTCHGYTITMTPNPYFLSCGEDGVTFVTRIKTSCIKEDCKDDILLNCRA
KTSVAICPIPYLLAVGSDSVRIYDRMLGTRATGNYAGTGTGMVARELPISLNN
ASRTALICSDGEOBILVSYSDYIYLPKDDTARELKTFSAEERLEKQPPVKR
LRLGDWPTGPRAPESPRDQSPNVSLQMSDMLSRWFEEASEVAQSNRGRG
SPRPGTQSQDISLTLPVSPDLEVSTAMEVDTFPAQFLOPSTSTSTMSAAHSTS
SPFSGTFLPLSSPDSQEQRSVEASGHHTHQSEFLRGPEITALLKRQLQRLKAE
QRRQELAAHTQQPSTSDQSSHEGSSQDPHAGY"
3875. .3992
/feature="L2 repeat: matches 1292. .1395 of consensus"
3993. .4316
/feature="AluY repeat: matches 4. .311 of consensus"
4317. .4657
/feature="L2 repeat: matches 1395. .1804 of consensus"
complement(4658. .4961)
/feature="match: GSS: Em:AQ488382.1"
/evidence="not_experimental"

```

```

repeat_region
/feature="L2 repeat: matches 2323. .2621 of consensus"
5417. .5702
/feature="AluX repeat: matches 1. .288 of consensus"
5710. .5804
/feature="HAI1 repeat: matches 734. .826 of consensus"
5810. .6092
/feature="AluX repeat: matches 5. .301 of consensus"
6245. .6554
/feature="AluY repeat: matches 2. .311 of consensus"
6571. .7311
/feature="L1M1 repeat: matches 5415. .6318 of consensus"
7319. .7583
/feature="L1M4 repeat: matches 4053. .4324 of consensus"
7584. .7869
/feature="AluJo repeat: matches 1. .281 of consensus"
7870. .8013
/feature="L1M4 repeat: matches 3912. .4053 of consensus"
8010. .8842
/feature="L1M4 repeat: matches 2663. .3517 of consensus"
8917. .10364
/feature="L1M7 repeat: matches 4735. .6165 of consensus"
10529. .12088
/feature="L1M1 repeat: matches 4735. .6316 of consensus"
12098. .12385
/feature="AluSg repeat: matches 21. .303 of consensus"
12386. .12422
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ081840.1"
/evidence="not_experimental"
12386. .12420
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ265659"
/evidence="not_experimental"
complement(12387. .12420)
/feature="match: GSS: Em:AQ000534
match: GSS: Em:AQ001615"
/evidence="not_experimental"
12426. .12589
/feature="L1M4 repeat: matches 2686. .2851 of consensus"
12590. .13036
/feature="L1M1 repeat: matches 4265. .4711 of consensus"
13037. .13344
/feature="AluY repeat: matches 1. .309 of consensus"
13345. .13542
/feature="L1M1 repeat: matches 4068. .4265 of consensus"
13546. .13630
/feature="trigger4(2ombi) repeat: matches 1. .86 of consensus"
13631. .13646
/feature="MER46A repeat: matches 44. .146 of consensus"
13650. .14477
/feature="L1MEC repeat: matches 2178. .3011 of consensus"
14478. .14778
/feature="AluX repeat: matches 1. .302 of consensus"
14783. .15077
/feature="AluSg repeat: matches 2. .293 of consensus"
15078. .15206
/feature="L1MEC repeat: matches 2403. .2179 of consensus"
15230. .15530
/feature="AluX repeat: matches 3. .312 of consensus"
15837. .17435
/feature="L1M4 repeat: matches 581. .2130 of consensus"
17436. .17738
/feature="AluX repeat: matches 1. .296 of consensus"
17739. .18061
/feature="L1M4 repeat: matches 248. .581 of consensus"
18086. .18306
/feature="MER30 repeat: matches 2. .230 of consensus"
18307. .18528
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ083937"
/evidence="not_experimental"
19519. .19834
repeat_region

```

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA

Query Match	9.4%;	Score 242.2;	DB 11;	Length 251;	
Best Local Similarity	98.8%;	Pred: No. 9.4e-55;			
Matches 244;	Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY	1125	AAGTCAATCAGATATTTCAACTCTTCTTACGCTCCCATCAAGTCTGTGATTGGGAAGTGAG	1184
Db	251	AAGTCAATCAGATATTTCAACTCTTCTTACGCTCCCATCAAGTCTGTGATTGGGAAGTGAG	192
QY	1185	TGAAACTGCAATGGAAGTAGATATCTCCAGCTGGAACAATTTCTTCAGCGCTTTACATCCTC	1244
Db	191	TGAAACTGCAATGGAAGTAGATATCTCCAGCTGGAACAATTTCTTCAGCGCTTTACATCCTC	132
QY	1245	TACAACTGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATCTTACTCC	1304
Db	131	TACAACTGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATCTTACTCC	72
QY	1305	TTTGCTATCTTCTCCAGACAGTGAACAAGCGACTGTTTGAGGCAATCTGGACACCAACAC	1364
Db	71	TTTGCTATCTTCTCCAGACAGTGAACAAGCGACTGTTTGAGGCAATCTGGACACCAACAC	12
QY	1365	ACATCAT 1371	
Db	11	ACACCT 5	

RESULT 21	AC069147	158324 bp	DNA	linear	HTG 01-SEP-2000
LOCUS	AC069147				
DEFINITION	Homo sapiens chromosome 1 clone RP11-105H9, WORKING DRAFT SEQUENCE,				
ACCESSION	AC069147				
VERSION	AC069147.3	GI:8099949			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				

SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 158324)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 158324)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-MAY-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On May 28, 2000 this sequence version replaced gi:8028173.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0105H09
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Chemistry: Dye-terminator; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 154865 bases at least Q40
 Consensus quality: 155623 bases at least Q30
 Consensus quality: 156154 bases at least Q20
 Insert size: 174000; agarose-Pf
 Insert size: 157724; sum-of-contigs
 Quality coverage: 7.55 in Q20 bases; agarose-pf
 Quality coverage: 8.39 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6587: contig of 6587 bp in length
 6588: gap of unknown length
 14247: contig of 7560 bp in length
 14348: gap of unknown length
 28851: contig of 14503 bp in length
 28852: gap of unknown length
 49498: contig of 20548 bp in length
 49499: gap of unknown length
 69546: contig of 19948 bp in length
 69547: gap of unknown length
 105850: contig of 36204 bp in length
 105851: gap of unknown length
 105951: contig of 52374 bp in length.

FEATURES
 source
 1. 158324
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-105H9"

misc_feature
 1. 6587
 /note="assembly_name:Contig15"
 misc_feature
 6588..14247
 /note="assembly_name:Contig16"
 misc_feature
 14348..28850
 /note="assembly_name:Contig17"
 clone_end:77
 vector_side:right
 28951..49496
 /note="assembly_name:Contig18"

misc_feature 49599..69546
 /note="assembly_name:Contig19"
 misc_feature 6947..105850
 /note="assembly_name:Contig20"
 misc_feature 105951..158324
 /note="assembly_name:Contig21"
 clone_end:SP6
 vector_side:right
 BASE COUNT 47660 a 30780 c 31018 g 48209 t 657 others
 ORIGIN
 Query Match 8.4%; Score 217.4; DB 2; Length 158324;
 Query Local Similarity 99.5%; Pred. No. 9.9e-48;
 Matches 218; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 687 AGGGATTATGACGAGGCGAGGACTACTGGAATGGTGGCCGTTTATTCCTTCCCATCT 746
 DB 25723 AGGGAATTATGACGAGGCGAGGACTACTGGAATGGTGGCCGTTTATTCCTTCCCATCT 25782
 QY 747 TAATAATAAGTCTGTCAGAGTGCATCTCTGTTCACAGTGAAGATGGTCAAGAGATTCT 806
 DB 25783 TAATAATAAGTCTGTCAGAGTGCATCTCTGTTCACAGTGAAGATGGTCAAGAGATTCT 25842
 QY 807 CGTTAGTACTCTTCAGATACATATATCTTTTACCCGAAAGATGATACACACGAGA 866
 DB 25843 CGTTAGTACTCTTCAGATACATATATCTTTTACCCGAAAGATGATACACACGAGA 25902
 QY 867 ACTTAAATCCCTTCGCGAAGAGAGAGAGAGT 905
 DB 25903 ACTTAAATCCCTTCGCGAAGAGAGAGAGAGT 25941

RESULT 22
 AC026140/c

LOCUS Homo sapiens chromosome 1 clone RP11-300C7 map 1, WORKING DRAFT
 DEFINITION AC026140 163230 bp DNA linear HTG 26-MAY-2000
 SEQUENCE, 17 unordered pieces.

ACCESSION AC026140.4 GI:8077102
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 1 (bases 1 to 163230)
 Homo sapiens chromosome 1, clone RP11-300C7

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 AUTHORS
 2 (bases 1 to 163230)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
 Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepey,K., Colangelo,M., Collins,S.,
 Collumore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferrel,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McKernan,K., McKernan,K., McPheeters,R.,
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
 Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On May 25, 2000 this sequence version replaced gi:7523800.
 All repeats were identified using RepeatMasker.
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L8103
 Center clone name: 300_C_7

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 152621 bases at least Q40
 Consensus quality: 157889 bases at least Q30
 Consensus quality: 159998 bases at least Q20
 Insert size: 157000; agarose-ff
 Insert size: 161630; sum-of-contigs
 Quality coverage: 4.9 in Q20 bases; agarose-ff
 Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence,
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1941: contig of 1941 bp in length
2      1942 2041: gap of 100 bp
3      2042 3714: contig of 1673 bp in length
4      3715 3814: gap of 100 bp
5      3815 7590: contig of 3776 bp in length
6      7591 7690: gap of 100 bp
7      7691 11450: contig of 3760 bp in length
8      11451 11550: gap of 100 bp
9      11551 18921: contig of 7371 bp in length
10     18922 19021: gap of 100 bp
11     19022 25215: contig of 6194 bp in length
12     25216 25315: gap of 100 bp
13     25316 30271: contig of 4956 bp in length
14     30272 30371: gap of 100 bp
15     30372 38111: contig of 7740 bp in length
16     38112 38211: gap of 100 bp
17     38212 47808: contig of 9597 bp in length
18     47809 47908: gap of 100 bp
19     47909 58047: contig of 10139 bp in length
20     58048 58147: gap of 100 bp
21     58148 71836: contig of 13689 bp in length
22     71837 71936: gap of 100 bp
23     71937 81124: contig of 9188 bp in length
24     81125 81224: gap of 100 bp
25     81225 92003: contig of 10778 bp in length
26     92003 92102: gap of 100 bp
27     92103 105933: contig of 13831 bp in length
28     105934 106033: gap of 100 bp
29     106034 121911: contig of 15878 bp in length
30     121912 122011: gap of 100 bp
31     122012 141942: contig of 19931 bp in length
32     141943 142042: gap of 100 bp
33     142043 163230: contig of 21188 bp in length.
      Location/Qualifiers
      1. .163230
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /map="1"
        /clone="RP11-300C7"
        /clone_lib="RPC1-11 Human Male BAC"

```

FEATURES
source

```

misc_feature      1. .1941
                  /note="assembly_fragment"
misc_feature      2042. .3714
                  /note="assembly_fragment"
                  clone_end:77
                  vector_side:right
misc_feature      3815. .7590
                  /note="assembly_fragment"
misc_feature      7691. .11450
                  /note="assembly_fragment"
misc_feature      11551. .18921
                  /note="assembly_fragment"
misc_feature      19022. .25215
                  /note="assembly_fragment"
misc_feature      25316. .30271
                  /note="assembly_fragment"
misc_feature      30372. .38111
                  /note="assembly_fragment"
misc_feature      38212. .47808
                  /note="assembly_fragment"
misc_feature      47909. .58047
                  /note="assembly_fragment"
misc_feature      58148. .71836
                  /note="assembly_fragment"
misc_feature      71937. .81124
                  /note="assembly_fragment"
misc_feature      81225. .92002
                  /note="assembly_fragment"
misc_feature      92103. .105933
                  /note="assembly_fragment"
misc_feature      106034. .121911
                  /note="assembly_fragment"
misc_feature      122012. .141942
                  /note="assembly_fragment"
misc_feature      142043. .163230
                  /note="assembly_fragment"
                  clone_end:SP6
                  vector_side:right

BASE COUNT      52925 a 30771 c 29634 g 48296 t 1604 others
ORIGIN
Query Match      8.4%; Score 217.4; DB 2; Length 163230;
Best Local Similarity 99.5%; Pred. No. 9.9e-48;
Matches 218; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 687 AGGGAATTATGAGGTCGAGGACTACTGGAATGGTTGCCGTTTATTCCTCCCATCT 746
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97524 AGGGAATTATGAGGTCGAGGACTACTGGAATGGTTGCCGTTTATTCCTCCCATCT 7465
QY 747 TAATAAAGTCTCGAGAGTGCATCTCTGTGTACATGAGATGGTCAAGAGATTCT 806
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97464 TAATAAAGTCTCGAGAGTGCATCTCTGTGTACATGAGATGGTCAAGAGATTCT 97405
QY 807 CGTTAGTTACTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGA 866
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97404 CTTTACTTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGA 97345
QY 867 ACTTAAACTCTCTTCGCGAGAGAGAGAGAGAGAGTT 905
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 97344 ACTTAAACTCTCTTCGCGAGAGAGAGAGAGAGTT 97306

RESULT 23
AC130131/c
LOCUS
DEFINITION      AC130131
                  Rattus norvegicus clone CH230-112N9, *** SEQUENCING IN PROGRESS
                  ***, 60 unordered pieces.
ACCESSION      AC130131
VERSION        AC130131.1 GI:22138359
KEYWORDS       HTG; HTGS_PHASE1.
SOURCE         Norway rat.
ORGANISM       Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1152: contig of 1152 bp in length
 1153: gap of unknown length
 1253: contig of 1117 bp in length
 2369: gap of unknown length
 2469: contig of 1151 bp in length
 3620: gap of unknown length
 3720: contig of 1171 bp in length
 4891: contig of 1171 bp in length
 4991: gap of unknown length
 6504: contig of 1513 bp in length
 6604: gap of unknown length
 7660: contig of 1056 bp in length
 7660: gap of unknown length
 9156: contig of 1396 bp in length
 9256: gap of unknown length
 10289: contig of 1033 bp in length
 10389: gap of unknown length
 11437: contig of 1048 bp in length
 11537: gap of unknown length
 13109: contig of 1572 bp in length
 13209: gap of unknown length
 14230: contig of 1021 bp in length
 14330: gap of unknown length
 14331: contig of 1321 bp in length
 15551: gap of unknown length
 17551: contig of 2147 bp in length
 17898: gap of unknown length
 17989: gap of unknown length
 19308: contig of 1310 bp in length
 19408: gap of unknown length
 20992: contig of 1584 bp in length
 21092: gap of unknown length
 22726: contig of 1634 bp in length
 22826: gap of unknown length
 23869: contig of 1043 bp in length
 23969: gap of unknown length
 25600: contig of 1631 bp in length
 25600: gap of unknown length
 27123: contig of 1423 bp in length
 27223: gap of unknown length
 28233: contig of 1010 bp in length
 28333: gap of unknown length
 29780: contig of 1447 bp in length
 29880: gap of unknown length
 31396: contig of 1516 bp in length
 31496: gap of unknown length
 32754: contig of 1258 bp in length
 32854: gap of unknown length
 35174: contig of 2320 bp in length
 35274: gap of unknown length
 37272: contig of 1998 bp in length
 37372: gap of unknown length
 39148: contig of 1776 bp in length
 39248: gap of unknown length
 40423: contig of 1175 bp in length
 40523: gap of unknown length
 41917: contig of 1394 bp in length
 42017: gap of unknown length
 45246: contig of 3229 bp in length
 45346: gap of unknown length
 47170: contig of 1824 bp in length
 47270: gap of unknown length
 49917: contig of 2647 bp in length
 50017: gap of unknown length
 51247: contig of 1230 bp in length
 51347: gap of unknown length
 54015: contig of 2668 bp in length
 54115: gap of unknown length
 56300: contig of 2185 bp in length
 56400: gap of unknown length
 58739: contig of 2339 bp in length

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.
 1 (bases 1 to 168307)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Ayale, M., Banks, T.,
 Barberia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, M., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, K., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delanay, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haviak, P., Hayes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.E.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louseghed, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, K., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwono, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,
 Rivers, M., Rojas, A., Rojokan, I., Roife, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sison, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Syatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 168307)
 Worley, K.C.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GH20
 Center clone name: CH230-112N9
 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 116575 bases at least Q40
 Consensus quality: 125466 bases at least Q30
 Consensus quality: 132182 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: LJ9554
 Center clone name: 294_I-9
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 172079 bases at least Q40
 Consensus quality: 175437 bases at least Q30
 Consensus quality: 176346 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 176969; sum-of-contigs
 Quality coverage: 8.0 in Q20 bases; agarose-fp
 Quality coverage: 8.0 in Q20 bases; sum-of-contigs

*** NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 772: contig of 772 bp in length
 773 872: gap of 100 bp
 873 2734: contig of 1862 bp in length
 2735 2834: gap of 100 bp
 2835 4210: contig of 1376 bp in length
 4211 4310: gap of 100 bp
 4311 5637: contig of 1327 bp in length
 5638 5737: gap of 100 bp
 5738 7170: contig of 1433 bp in length
 7171 7270: gap of 100 bp
 7271 8657: contig of 1387 bp in length
 8658 8757: gap of 100 bp
 8758 10043: contig of 1286 bp in length
 10044 10143: gap of 100 bp
 10144 11873: contig of 1730 bp in length
 11874 11973: gap of 100 bp
 11974 13309: contig of 1336 bp in length
 13310 13409: gap of 100 bp
 13410 14723: contig of 1314 bp in length
 14724 14823: gap of 100 bp
 14824 16735: contig of 1912 bp in length
 16736 16835: gap of 100 bp
 16836 19688: contig of 2853 bp in length
 19689 19788: gap of 100 bp
 19789 22930: contig of 3142 bp in length
 22931 23030: gap of 100 bp
 23031 25546: contig of 2516 bp in length
 25547 25646: gap of 100 bp
 25647 28158: contig of 2512 bp in length
 28159 28258: gap of 100 bp
 28259 32365: contig of 4107 bp in length
 32366 32465: gap of 100 bp
 32466 36376: contig of 3911 bp in length
 36377 36476: gap of 100 bp
 36477 40247: contig of 3771 bp in length
 40248 40347: gap of 100 bp
 40348 82256: contig of 41909 bp in length
 82257 82356: gap of 100 bp
 82357 92140: contig of 9784 bp in length
 92141 92240: gap of 100 bp
 92241 108763: contig of 16523 bp in length
 108764 108863: gap of 100 bp
 108864 141973: contig of 33110 bp in length

* 141974 142073: gap of 100 bp
 * 142074 170270: contig of 28197 bp in length
 * 170271 170270: gap of 100 bp
 * 170371 179269: contig of 8899 bp in length.
 FEATURES
 Source
 1. 179269
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-29419"
 /clone_lib="RPCI-23 Female Mouse BAC"
 1. 772
 /note="assembly_fragment"
 873. 2734
 /note="assembly_fragment"
 2835. 4210
 /note="assembly_fragment"
 4311. 5637
 /note="assembly_fragment"
 5738. 7170
 /note="assembly_fragment"
 7271. 8657
 /note="assembly_fragment"
 8758. 10043
 /note="assembly_fragment"
 10144. 11873
 /note="assembly_fragment"
 11974. 13309
 /note="assembly_fragment"
 13410. 14723
 /note="assembly_fragment"
 14824. 16735
 /note="assembly_fragment"
 16836. 19688
 /note="assembly_fragment"
 19789. 22930
 /note="assembly_fragment"
 23031. 25546
 /note="assembly_fragment"
 25647. 28158
 /note="assembly_fragment"
 28259. 32365
 /note="assembly_fragment"
 32466. 36376
 /note="assembly_fragment"
 36477. 40247
 /note="assembly_fragment"
 40348. 82256
 /note="assembly_fragment"
 82357. 92140
 /note="assembly_fragment"
 92241. 108763
 /note="assembly_fragment"
 108864. 141973
 /note="assembly_fragment"
 142074. 170270
 /note="assembly_fragment"
 170371. 179269
 /note="assembly_fragment"

Query Match 7.1%; Score 182.6; DB 2; Length 179269;
 Best Local Similarity 82.6%; Pred. No. 3.2e-38;
 Matches 209; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 687 AGGGAATATGAGGTCGAGGAGGACTGGAATGGTGGCCGCTTTTATTCCTTCCCATCT 746
 Db 46945 AGGGAATATGAGGTCGAGGAGGACTGGAATGGTGGCCGCTTTTATTCCTTCCCATCT 47004
 QY 747 TAATAATAAGTCTCGAGAGTGACATCTCTGTGTACAGTGAAGATGGTCAAGAGTTCT 806
 Db 47005 TAGTACAAATCATGACAGTGCATCATCTGTTACAGTGAAGATGGTCAAGAGTTCT 47064
 QY 807 CGTAGTACTTTCAGATTACATATATCTTTTGGACCCGAAAGATGATACAGCAGAGA 866

Db 47065 TGTCAGTATTCTTCAGATACATCTATCTTTTGACCCCAAGATGATCTGCACGAGA 47124
 QY 867 ACTTAAACTCTTCTGCGGAGAGAGAGAGAGAGTGGACAAACACCAAGTTAAAGCG 926
 Db 47125 ACTTAAACTCTTCTGCGGAGAGAGAGAGAGTGGACAAACACCAAGTTAAAGCG 926
 QY 927 TTTCAGATCTGCT 939
 Db 47185 CTTTCTCTTCTAGT 47197

RESULT 25
 AC127803/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-16L12, *** SEQUENCING IN PROGRESS
 AC127803
 VERSION 1
 KEYWORDS HTG; HTGS_PHASE1
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
 AUTHORS
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbarella,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cok,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Hale,S., Hamilton,K., Gorrell,J.H., Guevara,W., Gunaratne,P., Hales,N., Gill,R., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Khandekar,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C., Kravicic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Okwuonu,G., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Scherier,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

2 (bases 1 to 100694)
 Worley,K.C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXWC
 Center clone name: CH230-16L12
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 37710 bases at least Q40
 Consensus quality: 40353 bases at least Q30
 Consensus quality: 42652 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1772: contig of 1772 bp in length
 * 1773 1872: gap of unknown length
 * 1873 3356: contig of 1484 bp in length
 * 3357 3456: gap of unknown length
 * 3457 4772: contig of 1316 bp in length
 * 4773 4872: gap of unknown length
 * 4873 6219: contig of 1347 bp in length
 * 6220 6319: gap of unknown length
 * 6320 7626: contig of 1307 bp in length
 * 7627 7727: gap of unknown length
 * 7727 8787: contig of 1061 bp in length
 * 8788 8887: gap of unknown length
 * 8888 9962: contig of 1075 bp in length
 * 9963 10062: gap of unknown length
 * 10063 11362: contig of 1300 bp in length
 * 11363 11462: gap of unknown length
 * 11463 12825: contig of 1362 bp in length
 * 12825 12925: gap of unknown length
 * 12925 14287: contig of 1362 bp in length
 * 14287 14387: gap of unknown length
 * 14387 15370: contig of 1183 bp in length
 * 15370 15669: gap of unknown length
 * 15670 16828: contig of 1159 bp in length
 * 16829 16929: gap of unknown length
 * 16929 18383: contig of 1455 bp in length
 * 18384 18483: gap of unknown length
 * 18484 20335: contig of 1851 bp in length
 * 20335 20435: gap of unknown length
 * 20435 21570: contig of 1135 bp in length
 * 21570 21669: gap of unknown length
 * 21670 23401: contig of 1731 bp in length
 * 23401 23501: gap of unknown length
 * 23501 25107: contig of 1607 bp in length
 * 25107 25207: gap of unknown length
 * 25207 25208: contig of 1281 bp in length
 * 25208 26488: gap of unknown length
 * 26489 26589: contig of 1494 bp in length
 * 26589 28082: gap of unknown length
 * 28082 28182: contig of 1665 bp in length
 * 28183 29847: gap of unknown length
 * 29848 29948: contig of 1980 bp in length
 * 29948 31927: gap of unknown length
 * 31927 32027: gap of unknown length
 * 32027 33704: contig of 1677 bp in length
 * 33704 33804: gap of unknown length
 * 33804 35309: contig of 1505 bp in length
 * 35309 35409: gap of unknown length
 * 35410 36919: contig of 1510 bp in length

Db	Accession	Length	Score	E-value	Identical	Similarity	Conservative	Mismatches	Indels	Gaps
Db	36920	37019: gap of unknown length	6.5%	Score 166.6; DB 2: Length 100694;	0	0	0	0	0	0
Db	37020	38399: contig of 1380 bp in length	65.4%	Pred. No. 7.1e-34;	0	0	0	0	0	0
Db	38400	38499: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	38500	39683: contig of 1184 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	39684	39783: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	39784	40830: contig of 1047 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	40831	40930: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	40931	42811: contig of 1881 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	42812	42911: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	42912	44020: contig of 1109 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	44021	44120: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	44121	45757: contig of 1637 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	44578	45857: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	45858	47651: contig of 1794 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	47652	47751: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	49041	49040: contig of 1289 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	49141	49140: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	49141	51374: contig of 2234 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	51475	51474: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	53653	53652: contig of 2178 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	53753	53752: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	55443	55442: contig of 1690 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	55443	55442: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	57966	57965: contig of 2423 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	58066	58065: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	59662	59661: contig of 1596 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	59662	59761: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	62023	62022: contig of 2261 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	62123	62122: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	64714	64713: contig of 2591 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	67919	67918: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	68019	71123: contig of 3105 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	71124	71123: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	73931	73930: contig of 2707 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	74031	74030: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	76005	76004: contig of 1974 bp in length	0	Mismatches 244;	0	0	0	0	0	0
Db	76105	76104: gap of unknown length	0	Indels 129;	0	0	0	0	0	0
Db	77245	77244: contig of 1040 bp in length	0	Mismatches 244;						

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 179269)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhalil, B., Brown, A., Camarata, J., Campionato, A., Chang, J.,
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faroo, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
MacLean, C., Macdonald, P., Major, J., McKernan, K., Meldrum, J., Meneus, L.,
McCarthy, M., McEwan, P., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Mihova, I., Mlenga, V., Morphy, T., O'Connor, T., O'Donnell, P., O'Neill, D.,
Norbu, C., Norman, C.H., Phunkhang, P., Pierre, N., Pollara, V.,
Oliver, J., Peterson, K., Phunkhang, P., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Raymond, C., Retta, R., Rieback, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Roman, J., Rosetti, M., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Stoian, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J.,
Strauss, N., Subramanian, A., Talamas, J., Teafaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemlek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 28, 2002 this sequence version replaced gi:19745074.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19554
Center clone name: 294_L19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 172079 bases at least Q40
Consensus quality: 175437 bases at least Q30
Consensus quality: 176346 bases at least Q20
Insert size: 176969; agarose-
Insert size: 176969; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-
Quality coverage: 8.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 772: contig of 772 bp in length
* 773 872: gap of 100 bp
* 873 873: contig of 1862 bp in length
* 873 2834: gap of 100 bp
* 2835 4210: contig of 1376 bp in length
* 4211 4310: gap of 100 bp
* 4311 5637: contig of 1327 bp in length
* 5638 5737: gap of 100 bp
* 5738 5737: gap of 100 bp

* 5738 7170: contig of 1433 bp in length
 * 7171 7270: gap of 100 bp
 * 7271 8657: contig of 1387 bp in length
 * 8658 8757: gap of 100 bp
 * 8758 10043: contig of 1286 bp in length
 * 10044 10143: gap of 100 bp
 * 10144 11873: contig of 1730 bp in length
 * 11874 11973: gap of 100 bp
 * 11974 13309: contig of 1336 bp in length
 * 13310 14723: contig of 1314 bp in length
 * 14724 14823: gap of 100 bp
 * 14824 16735: contig of 1912 bp in length
 * 16736 16835: gap of 100 bp
 * 16836 19688: contig of 2853 bp in length
 * 19689 19788: gap of 100 bp
 * 19789 22930: contig of 3142 bp in length
 * 22931 23030: gap of 100 bp
 * 23031 25546: contig of 2516 bp in length
 * 25547 25646: gap of 100 bp
 * 25647 28158: contig of 2512 bp in length
 * 28159 28258: gap of 100 bp
 * 28259 32365: contig of 4107 bp in length
 * 32366 32465: gap of 100 bp
 * 32466 36376: contig of 3911 bp in length
 * 36377 36476: gap of 100 bp
 * 36477 40247: contig of 3771 bp in length
 * 40248 40347: gap of 100 bp
 * 40348 82256: contig of 41909 bp in length
 * 82257 82356: gap of 100 bp
 * 82357 92140: contig of 9784 bp in length
 * 92141 92240: gap of 100 bp
 * 92241 108763: contig of 16523 bp in length
 * 108764 108863: gap of 100 bp
 * 108864 141973: contig of 33110 bp in length
 * 141974 142073: gap of 100 bp
 * 142074 170270: contig of 28197 bp in length
 * 170271 170370: gap of 100 bp
 * 170371 179269: contig of 8899 bp in length.

FEATURES

source

1. .179269
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RP23-29419"
 1. .772
 /note="assembly_fragment"
 873. .2734
 /note="assembly_fragment"
 2835. .4210
 /note="assembly_fragment"
 4311. .5637
 /note="assembly_fragment"
 5738. .7170
 /note="assembly_fragment"
 7271. .8657
 /note="assembly_fragment"
 8758. .10043
 /note="assembly_fragment"
 10144. .11873
 /note="assembly_fragment"
 11974. .13309
 /note="assembly_fragment"
 13410. .14723
 /note="assembly_fragment"
 14824. .16735
 /note="assembly_fragment"
 16836. .19688
 /note="assembly_fragment"
 19789. .22930
 /note="assembly_fragment"
 23031. .25546
 /note="assembly_fragment"

misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature

misc_feature 25647. .28158
 /note="assembly_fragment"
 misc_feature 28259. .32365
 /note="assembly_fragment"
 misc_feature 32466. .36376
 /note="assembly_fragment"
 misc_feature 36477. .40247
 /note="assembly_fragment"
 misc_feature 40348. .82256
 /note="assembly_fragment"
 misc_feature 82357. .92140
 /note="assembly_fragment"
 misc_feature 92241. .108763
 /note="assembly_fragment"
 misc_feature 108864. .141973
 /note="assembly_fragment"
 misc_feature 142074. .170270
 /note="assembly_fragment"
 misc_feature 170371. .179269
 /note="assembly_fragment"

Query Match

Best Local Similarity 6.2%; Score 159.6; DB 2; Length 179269;
 Matches 171; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 251 AGTTTTCACACAAATTCGTTCCAGGCGCAGCAACATATTTAGTGCAGAGTTCTTAC 310
 Db 174144 AGTTTTCACACAAATTCGTTCCAGGCGCAGCAACATATTTAGTGCAGAGTTCTTAC 310
 QY 311 CTTGTACAAATGATAACAGATTGTATCTGCTCTGGAGATGAGTAATATTTATACCA 370
 Db 174084 CGTGCACAGATGATAAGCAGATTGTCTGCTCTGGAGATGAGTAATATTTATACCA 370
 QY 371 ACGTTGACCAAGATGCAGAAACCAACAGACAAATTCGTTCTGTTATGGAACCTA 430
 Db 174024 ACGTTGACCAAGATGCAGAAACCAACAGACAAATTCGTTCTGTTATGGAACCTA 430
 QY 431 CTTATGAGAT 440
 Db 173964 CTTATGAGAT 173955

RESULT 28

AC127803

LOCUS

DEFINITION

AC127803

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 100694)

Muzdy,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,

Barbata,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinah,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,

Hollins,B.,

Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. F.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisege, H.,
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oriado, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojuekan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 100694)
 Worley, K. C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GYX
 Center clone name: CH230-16L12

 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 37710 bases at least Q40
 Consensus quality: 40353 bases at least Q30
 Consensus quality: 42652 bases at least Q20

 NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1772: contig of 1772 bp in length
 * 1773: gap of unknown length
 * 1873: contig of 1484 bp in length
 * 3356: gap of unknown length
 * 3456: gap of unknown length
 * 4772: contig of 1316 bp in length
 * 4773: gap of unknown length
 * 4773: contig of 1347 bp in length
 * 4873: gap of unknown length
 * 6219: gap of unknown length
 * 6320: contig of 1307 bp in length
 * 7627: gap of unknown length
 * 7727: contig of 1061 bp in length
 * 8887: gap of unknown length
 * 8888: contig of 1075 bp in length
 * 9363: contig of 10062 bp in length
 * 10063: contig of 1300 bp in length
 * 11362: contig of 1300 bp in length
 * 11363: gap of unknown length
 *
 * 11463: contig of 1362 bp in length
 * 12924: gap of unknown length
 * 12925: contig of 1362 bp in length
 * 14287: gap of unknown length
 * 14386: contig of 1183 bp in length
 * 15569: gap of unknown length
 * 15669: contig of 1159 bp in length
 * 16928: gap of unknown length
 * 16929: contig of 1455 bp in length
 * 18384: gap of unknown length
 * 18483: contig of 1851 bp in length
 * 20334: gap of unknown length
 * 20434: contig of 1135 bp in length
 * 21569: gap of unknown length
 * 21669: contig of 1731 bp in length
 * 23400: gap of unknown length
 * 23500: contig of 1607 bp in length
 * 25107: gap of unknown length
 * 25207: contig of 1281 bp in length
 * 26488: gap of unknown length
 * 26589: contig of 1494 bp in length
 * 28082: gap of unknown length
 * 28182: contig of 1665 bp in length
 * 29847: gap of unknown length
 * 29848: contig of 1980 bp in length
 * 31927: gap of unknown length
 * 32027: contig of 1677 bp in length
 * 33704: gap of unknown length
 * 33804: contig of 1505 bp in length
 * 35309: gap of unknown length
 * 35409: contig of 1510 bp in length
 * 36919: gap of unknown length
 * 37019: contig of 1380 bp in length
 * 38399: gap of unknown length
 * 38499: contig of 1184 bp in length
 * 39683: gap of unknown length
 * 39783: contig of 1047 bp in length
 * 40830: gap of unknown length
 * 40930: contig of 1881 bp in length
 * 42811: gap of unknown length
 * 42911: contig of 1109 bp in length
 * 44020: gap of unknown length
 * 44120: contig of 1637 bp in length
 * 45757: gap of unknown length
 * 45857: contig of 1794 bp in length
 * 47651: gap of unknown length
 * 47751: contig of 1289 bp in length
 * 49040: gap of unknown length
 * 49140: contig of 2234 bp in length
 * 51374: gap of unknown length
 * 51474: contig of 2178 bp in length
 * 53652: gap of unknown length
 * 53752: contig of 1690 bp in length
 * 55442: gap of unknown length
 * 55542: contig of 2423 bp in length
 * 57965: gap of unknown length
 * 58065: contig of 1596 bp in length
 * 59661: gap of unknown length
 * 59761: contig of 2261 bp in length
 * 62022: gap of unknown length
 * 62122: contig of 2591 bp in length
 * 64713: gap of unknown length
 * 64813: contig of 3105 bp in length
 * 67918: gap of unknown length
 * 68018: contig of 3105 bp in length
 * 71123: gap of unknown length
 * 71223: contig of 2707 bp in length
 * 73930: gap of unknown length
 * 74030: contig of 1974 bp in length
 * 76004: gap of unknown length
 * 76104: contig of 1040 bp in length
 * 77144: gap of unknown length
 * 77244: contig of 2830 bp in length
 * 80074: gap of unknown length
 * 77245: contig of 2830 bp in length

* 80075 80174: gap of unknown length
 * 80175 83578: contig of 3404 bp in length
 * 83579 83678: gap of unknown length
 * 83679 86140: contig of 2462 bp in length
 * 86141 86240: gap of unknown length
 * 86241 89933: contig of 3693 bp in length
 * 89934 90034: gap of unknown length
 * 90034 94841: contig of 4807 bp in length
 * 94841 94940: gap of unknown length
 * 94941 100694: contig of 5754 bp in length.

Location/Qualifiers
 1. 100694
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-16L12"

BASE COUNT 25190 a 18744 c 18830 g 24955 t 12975 others

ORIGIN

Query Match 5.5%; Score 142.4; DB 2; Length 100694;
 Best Local Similarity 93.1%; Pred. No. 2.9e-27;
 Matches 149; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2166 GATAAAGAGCCAAATTCCTGGGGTCTAACTTTGTAATGAGTGGTTCTGACTGTGGCCA 2225
 |||||
 Db 76852 GATAAAGAGCCAAATTCCTGGGGTCTAACTTTGTAATGAGTGGTTCTGACTGTGGCCA 76911
 |||||

QY 2226 CATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 2285
 |||||
 Db 76912 TATCTTCATCTGGGACCGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 76971
 |||||

QY 2286 TCTGGTAACTGCTGCGACCCACATCGGTTTGACCAAT 2325
 |||||
 Db 76972 CGTGTCAACTGCTGCGACCCACATTCATTCACCAAGT 77011
 |||||

RESULT 29
 AC100993/c
 LOCUS
 DEFINITION Mus musculus clone RP23-77G13, LOW-PASS SEQUENCE SAMPLING.
 AC100993
 AC100993.1 GI:17059767
 VERSION
 KEYWORDS HTG; HTGS PHASED.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 59143)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-77G13
 Unpublished
 2 (bases 1 to 59143)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
 Hago, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
 Lamazares, K., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrum, J.,
 Meneus, L., Milhova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

TITLE JOURNAL COMMENT

Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WtBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L14696
 Center clone name: 77_G_13

* NOTE: This record contains 74 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 712: contig of 712 bp in length
 813 812: gap of 100 bp
 813 1502: contig of 690 bp in length
 1503 1602: gap of 100 bp
 1603 2325: contig of 723 bp in length
 2326 2425: gap of 100 bp
 2426 3145: contig of 720 bp in length
 3146 3245: gap of 100 bp
 3246 3906: contig of 661 bp in length
 3907 4006: gap of 100 bp
 4007 4708: contig of 702 bp in length
 4709 4808: gap of 100 bp
 4809 5483: contig of 675 bp in length
 5484 5583: gap of 100 bp
 5584 6277: contig of 694 bp in length
 6278 6377: gap of 100 bp
 6378 7081: contig of 704 bp in length
 7082 7181: gap of 100 bp
 7182 7899: contig of 718 bp in length
 7900 7999: gap of 100 bp
 8000 8689: contig of 690 bp in length
 8690 8789: gap of 100 bp
 8790 9517: contig of 728 bp in length
 9518 9617: gap of 100 bp
 9618 10249: contig of 632 bp in length
 10250 10349: gap of 100 bp
 10350 11057: contig of 708 bp in length
 11058 11157: gap of 100 bp
 11158 11850: contig of 693 bp in length
 11851 11950: gap of 100 bp
 12621 12621: contig of 671 bp in length
 12722 13417: contig of 696 bp in length
 13418 13517: gap of 100 bp
 13518 14208: contig of 691 bp in length
 14209 14308: gap of 100 bp
 14309 15001: contig of 693 bp in length
 15002 15101: gap of 100 bp
 15102 15795: contig of 694 bp in length
 15796 15895: gap of 100 bp
 15896 16595: contig of 700 bp in length
 16596 16695: gap of 100 bp
 16696 17432: contig of 737 bp in length
 17433 17532: gap of 100 bp
 17533 18233: contig of 701 bp in length
 18234 18333: gap of 100 bp
 18334 19056: contig of 723 bp in length

* 47919 48611: contig of 693 bp in length
 * 48612 48711: gap of 100 bp
 * 48712 49415: contig of 704 bp in length
 * 49416 49515: gap of 100 bp
 * 49516 50231: contig of 716 bp in length
 * 50232 50331: gap of 100 bp
 * 50332 51060: contig of 729 bp in length
 * 51061 51160: gap of 100 bp
 * 51161 51899: contig of 739 bp in length
 * 51900 51999: gap of 100 bp
 * 52000 52719: contig of 720 bp in length
 * 52720 52819: gap of 100 bp
 * 52820 53540: contig of 721 bp in length
 * 53541 53640: gap of 100 bp
 * 53641 54332: contig of 692 bp in length
 * 54333 54432: gap of 100 bp
 * 54433 55139: contig of 707 bp in length
 * 55140 55239: gap of 100 bp

Query Match 4.9%; Score 125.2; DB 2; Length 59143;
 Best Local Similarity 70.4%; Pred. No. 1.4e-22;
 Matches 197; Conservative 0; Mismatches 78; Indels 5; Gaps 2;

QY 1 ATGCTCGGGTGGCTCTACCCACACCTGTGTGGAGCTGAGGAAAGTCCCTCGGG 60
 DB 1138 AAGTTTATAGTATGTTCTGCTGCACCTGTATAGGACATGAGGAAAGAACCTCAGG 1079
 QY 61 CTGGAGGACCCCTCCCGGCTGGAGTCTGCTACCTGGGAGAGAGATTTATCCAAAGA 120
 DB 1078 CTACAAAGTCTCTCCCTCTGAGAGCCACTGCTG---AGAAGAGACTTAGGCAAAACA 1022
 QY 121 TTAACACTTGAAGCAACCTCTAATGTGCATGATGGTGTGTTAATACATCTGTTGGAAT 180
 DB 1021 CTAAGCCTTCAGGCAACATGAATGTTCATAAGCAGGTAGTAGT--GATCTTTTAGAAG 964
 QY 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAACTAGTAACTCT 240
 DB 963 GACACTAGCAATATATTTTGTGCTTATAGTATGATGCAAGTATAGTAACTAGTAACTCT 904
 QY 241 TACAGCAAGAGTTTGTGACAACTTCGTTTCAGGCAACC 280
 DB 903 CACAGCAGAAAGGTTTGTGACAACTTCGTTTCAGGCAACC 864

RESULT 30
 AL732543 232993 bp DNA linear HTG 17-AUG-2002
 LOCUS Mus musculus chromosome 4 clone RP23-13D22, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 5 unordered pieces.

ACCESSION AL732543
 VERSION AL732543.2 GI:22415946
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 232993)
 AUTHORS Lovell, J.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Aug 21, 2002 this sequence version replaced gi:21213781.

COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: bM13D22
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads

* 19057 19156: gap of 100 bp
 * 19157 19852: contig of 696 bp in length
 * 19853 19952: gap of 100 bp
 * 19953 20667: contig of 715 bp in length
 * 20668 20767: gap of 100 bp
 * 20768 21449: contig of 682 bp in length
 * 21450 21549: gap of 100 bp
 * 21550 22161: contig of 612 bp in length
 * 22162 22661: gap of 100 bp
 * 22662 22957: contig of 696 bp in length
 * 22958 23057: gap of 100 bp
 * 23058 23779: contig of 722 bp in length
 * 23780 23879: gap of 100 bp
 * 23880 24591: contig of 712 bp in length
 * 24592 24691: gap of 100 bp
 * 24692 25418: contig of 727 bp in length
 * 25419 25518: gap of 100 bp
 * 25519 26226: contig of 708 bp in length
 * 26227 26326: gap of 100 bp
 * 26327 27019: contig of 693 bp in length
 * 27020 27119: gap of 100 bp
 * 27120 27814: contig of 695 bp in length
 * 27815 27914: gap of 100 bp
 * 27915 28612: contig of 698 bp in length
 * 28613 28712: gap of 100 bp
 * 28713 29436: contig of 724 bp in length
 * 29437 29536: gap of 100 bp
 * 29537 30238: contig of 702 bp in length
 * 30239 30338: gap of 100 bp
 * 30339 31043: contig of 705 bp in length
 * 31044 31143: gap of 100 bp
 * 31144 31881: contig of 738 bp in length
 * 31882 31981: gap of 100 bp
 * 31982 32694: contig of 713 bp in length
 * 32695 32794: gap of 100 bp
 * 32795 33478: contig of 684 bp in length
 * 33479 33578: gap of 100 bp
 * 33579 34276: contig of 698 bp in length
 * 34277 34376: gap of 100 bp
 * 34377 35074: contig of 698 bp in length
 * 35075 35174: gap of 100 bp
 * 35175 35878: contig of 704 bp in length
 * 35879 35978: gap of 100 bp
 * 35979 36661: contig of 683 bp in length
 * 36662 36761: gap of 100 bp
 * 36762 37561: contig of 700 bp in length
 * 37562 38274: contig of 713 bp in length
 * 38275 38374: gap of 100 bp
 * 38375 39052: contig of 678 bp in length
 * 39053 39152: gap of 100 bp
 * 39153 39852: contig of 700 bp in length
 * 39853 39952: gap of 100 bp
 * 39953 40654: contig of 702 bp in length
 * 40655 40754: gap of 100 bp
 * 40755 41360: contig of 606 bp in length
 * 41361 41460: gap of 100 bp
 * 41461 42155: contig of 695 bp in length
 * 42156 42255: gap of 100 bp
 * 42256 42953: contig of 698 bp in length
 * 42954 43053: gap of 100 bp
 * 43054 43774: contig of 721 bp in length
 * 43775 43874: gap of 100 bp
 * 43875 44596: contig of 722 bp in length
 * 44597 44696: gap of 100 bp
 * 44697 45410: contig of 714 bp in length
 * 45411 45510: gap of 100 bp
 * 45511 46201: contig of 691 bp in length
 * 46202 46301: gap of 100 bp
 * 46302 47012: contig of 711 bp in length
 * 47013 47112: gap of 100 bp
 * 47113 47818: contig of 706 bp in length
 * 47819 47918: gap of 100 bp

Consensus quality: 231947 bases at least Q40
 Consensus quality: 232218 bases at least Q30
 Consensus quality: 232424 bases at least Q20
 Insert size: 232593; sum-of-contigs
 Insert size: 233521; 6.4% error; agarose-fp
 Quality coverage: 5.47x in Q20 bases; sum-of-contigs Quality
 coverage: 5.45x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 131776: contig of 131776 bp in length
 131777 131876: gap of 100 bp
 131877 180768: contig of 48892 bp in length
 180769 180868: gap of 100 bp
 180869 227965: contig of 47097 bp in length
 227966 228065: gap of 100 bp
 228066 230501: contig of 2436 bp in length
 230502 230601: gap of 100 bp
 230602 232993: contig of 2392 bp in length.

FEATURES

source

1. 232993
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-13D22"
 /clone_lib="RPCI-23"
 1. 131776
 /note="assembly_fragment:01399
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 131877. 180768
 /note="assembly_fragment:01212
 fragment_chain:1"
 180869. 227965
 /note="assembly_fragment:00644
 fragment_chain:1"
 228066. 230501
 /note="assembly_fragment:02215
 fragment_chain:2"
 230602. 232993
 /note="assembly_fragment:02207
 fragment_chain:2"

BASE COUNT 67150 a 50371 c 47633 g 67439 t 400 others
 ORIGIN

Query Match 4.9%; Score 125.2; DB 2; Length 232993;
 Best Local Similarity 70.4%; Pred. No. 1.6e-22;
 Matches 197; Conservative 0; Mismatches 78; Indels 5; Gaps 2;
 QY 1 ATGTCGGGGTGGCTCCTACCCACACCTGTTGTGGGACGTGAGGAAAGGCCCTCGG 60
 Db 46014 AAGTTTATAGTGAATTTCTGCTGCACCTGTTATAGGACATGAGGAAAGACCTCAGG 46073
 QY 61 CTGGAGGACCGTCGCGGCTCGGAGTCCCTACCTGGGAAGAGATTTATCCAAAGA 120
 Db 46074 CTACAACTCTGTCCTCCTGAAGGACCACTGCCTG---AGAAGAGACTTAGGCAAAACA 46130
 QY 121 TTAACACTTGAAGCAACCCCTTAATGTGTCATGATGTTGTGTTATATACATCTGTTGGAT 180
 Db 46131 CTAAGCCCTTCAGGCAACATGAATGTTATCAAGCAGGTAGTAGT--GATCTTTTAGAAG 46188
 QY 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATAGTAATAGTAATCCT 240
 Db 46189 GACACTGACAAATATATTTCTGCTGGCTAGATGCAAGTTAGTAGTAGTATGCT 46248
 QY 241 TACAGCAGAAAGGTTTTCACAAACAAATTCGTTACGGGCACC 280

Db 46249 CACAGCAGAAAGGCTTTCACAAACAAATCCTCAGGCACC 46288

RESULT 31
 AC069147/c
 LOCUS

DEFINITION

AC069147 158324 bp DNA linear HTG 01-SEP-2000
 Homo sapiens chromosome 1 clone RP11-105H9, WORKING DRAFT SEQUENCE,
 7 unordered pieces.

ACCESSION

AC069147

VERSION

AC069147.3

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens

REFERENCE

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Waterston,R.H.

Wed Mar 26 09:38:48 2003

FEATURES
source
1. .102258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q24"
/clone_lib="RPCI-1"
/clone="RP1-295C6"
891. .1180
/note="AluSq repeat: matches 298. .9 of consensus"
repeat_region
1375. .1683
/note="AluSp repeat: matches 303. .1 of consensus"
repeat_region
2028. .2310
/note="AluSq repeat: matches 300. .1 of consensus"
repeat_region
2612. .2903
/note="AluSx repeat: matches 301. .1 of consensus"
misc_feature
complement(4475. .4866)
/note="match: Z23618 STS containing (CA) repeat"
repeat_region
4625. .4670
/note="23 copies of GT 100 % conserved; differs from Z23618"
4766. .4897
/note="MIR2 repeat: matches 145. .1 of consensus"
repeat_region
4969. .5140
/note="MER42C repeat: matches 1475. .1300 of consensus"
repeat_region
5146. .5448
/note="AluJo repeat: matches 302. .1 of consensus"
repeat_region
5303. .5795
/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region
5797. .5925
/note="MER42C repeat: matches 1265. .1124 of consensus"
repeat_region
5923. .6039
/note="L1MB6 repeat: matches 920. .804 of consensus"
repeat_region
6123. .6415
/note="AluSx repeat: matches 1. .292 of consensus"
repeat_region
7799. .8099
/note="AluSx repeat: matches 302. .1 of consensus"
repeat_region
9002. .9302
/note="AluSp repeat: matches 2. .303 of consensus"
repeat_region
9692. .9983
/note="AluJo repeat: matches 1. .301 of consensus"
repeat_region
10986. .11071
/note="MIR2 repeat: matches 146. .60 of consensus"
repeat_region
11583. .11766
/note="MIR repeat: matches 262. .70 of consensus"
repeat_region
12180. .12479
/note="AluSg repeat: matches 1. .300 of consensus"
repeat_region
13246. .13500
/note="MIR repeat: matches 256. .1 of consensus"
repeat_region
13907. .13973
/note="MIR repeat: matches 75. .141 of consensus"
repeat_region
13975. .14489
/note="MER1A repeat: matches 526. .1 of consensus"
repeat_region
15702. .16064
/note="MER1C repeat: matches 1. .371 of consensus"
repeat_region
16144. .16585
/note="MIR1C repeat: matches 9. .466 of consensus"
repeat_region
17173. .17534
/note="LNR2 repeat: matches 105. .449 of consensus"
prim_transcript
<17786. .>18427
/note="match: multiple ESTs; match: AA165668 C16515 AA077391"
repeat_region
18004. .18120
/note="3 copies of 39 mer 81 % conserved"
prim_transcript
19377. .>21532
/note="match: multiple ESTs; match: N46026 H60052 H12822 T62974 AA283144; match: T57835 F00049 N75628 AA382351 AA300207; match: W37181 N52820 AA337499 W87891 H84729;

The true left end of clone 295C6 is at 1 in this sequence. The true right end of clone 295C6 is at 102258.
295C6 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.
For further details see <http://bacpac.med.buffalo.edu/>.

Query Match 3.8%; Score 98; DB 2; Length 158324;
Best Local Similarity 100.0%; Pred. No. 4e-15;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCGCGGGTGCTCCTACCCACACCTGTGTGGGACGTGAGGAAAGTCCCTCGGG 60
|||||
Db 93904 ATGTCGCGGGTGCTCCTACCCACACCTGTGTGGGACGTGAGGAAAGTCCCTCGGG 93845
|||||

QY 61 CTGAGGACCCGTCCTCCGCGAGTGGCTACCTCGGG 98
|||||
Db 93844 CTGAGGACCCGTCCTCCGCGAGTGGCTACCTCGGG 93807
|||||

RESULT 32
HS295C6 102258 bp DNA linear PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 295C6 on chromosome 1q24. Contains
DEFINITION ESTs, CA repeat, STS and CpG island.
ACCESSION 297876
VERSION 1q24; CpG island; repeat polymorphism.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Grafham,D.
Direct Submission
Submitted (27-OCT-1997) Chromosome 1 Project Group
(<http://www.sanger.ac.uk/HGP/Chr1/>) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquery@sanger.ac.uk
HumQuery@Sanger.ac.uk
On Nov 2, 1997 this sequence version replaced g1:2465042.
IMPORTANT: This sequence is the entire insert of clone 295C6.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre chromosome 1
mapping group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1/>
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

similar to endogenous retrovirus POL POLYPROTEIN"
 <22156..22928
 /note="match: multiple ESTs; match: AA401243 AA258918
 AA248892 AA096209 R36280"
 23025..23060
 /note="3 copies of 12 mer 94 % conserved"
 23519..23865
 /note="LTR2 repeat: matches 105..449 of consensus"
 24473..24785
 /note="AluSg repeat: matches 1..299 of consensus"
 25037..25058
 /note="16 copies of 2 mer 88 % conserved"
 25877..25987
 /note="L1P45 repeat: matches 788..892 of consensus"
 26022..26312
 /note="AluSp repeat: matches 1..295 of consensus"
 26752..27050
 /note="AluSg repeat: matches 1..299 of consensus"
 28012..28312
 /note="AluSg repeat: matches 1..303 of consensus"
 28316..29066
 /note="L1ME3A repeat: matches 109..907 of consensus"
 29517..29629
 /note="MIR repeat: matches 262..151 of consensus"
 29631..29928
 /note="AluY repeat: matches 1..295 of consensus"
 29937..30237
 /note="AluSg repeat: matches 1..301 of consensus"
 30238..30315
 /note="MIR repeat: matches 158..76 of consensus"
 31096..31287
 /note="MIR repeat: matches 56..262 of consensus"
 31289..31379
 /note="MIR2 repeat: matches 146..56 of consensus"
 31479..31525
 /note="MIR2 repeat: matches 146..95 of consensus"
 33069..33318
 /note="MIR repeat: matches 9..253 of consensus"
 35565..35744
 /note="MIR repeat: matches 262..74 of consensus"
 36165..36466
 /note="AluSg repeat: matches 1..302 of consensus"
 36692..36944
 /note="MIR repeat: matches 3..240 of consensus"
 36999..37299
 /note="AluSg repeat: matches 301..1 of consensus"
 37924..38114
 /note="L1MA2 repeat: matches 1055..866 of consensus"
 38108..38475
 /note="MSPA repeat: matches 426..1 of consensus"
 38479..38657
 /note="L1MA2 repeat: matches 874..687 of consensus"
 38658..38957
 /note="AluSg repeat: matches 300..1 of consensus"
 38958..39580
 /note="L1MA3 repeat: matches 699..85 of consensus"
 39585..39875
 /note="AluSg repeat: matches 292..1 of consensus"
 39877..39951
 /note="L1MA9 repeat: matches 88..14 of consensus"
 39944..40919
 /note="L1 repeat: matches 4416..5390 of consensus"
 40769..41662
 /note="L1P42 repeat: matches 1..893 of consensus"
 41764..42556
 /note="L1 repeat: matches 5133..4305 of consensus"
 42557..42857
 /note="AluY repeat: matches 1..301 of consensus"
 42876..43925
 /note="L1 repeat: matches 4313..3215 of consensus"
 44235..44498
 /note="MER43 repeat: matches 3..272 of consensus"
 44788..44860

repeat_region /note="MIR repeat: matches 154..82 of consensus"
 44933..45220
 /note="AluSg repeat: matches 289..2 of consensus"
 45304..45336
 repeat_region /note="AluY repeat: matches 1..301 of consensus"
 45337..45637
 repeat_region /note="AluY repeat: matches 1..301 of consensus"
 45934..46220
 repeat_region /note="AluY repeat: matches 302..14 of consensus"
 46817..46935
 repeat_region /note="MIR repeat: matches 35..154 of consensus"
 47393..47677
 repeat_region /note="AluSg repeat: matches 1..302 of consensus"
 48309..48404
 repeat_region /note="MIR repeat: matches 48..140 of consensus"
 48620..48862
 repeat_region /note="MER21B repeat: matches 347..102 of consensus"
 48865..49159
 repeat_region /note="AluSc repeat: matches 1..296 of consensus"
 51182..51467

Query Match 3.7%; Score 96.4; DB 9; Length 102258;
 Best Local Similarity 99.0%; Pred. No. 1.1e-14;
 Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTCCTGGGTGCTCTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 60
 |||||
 Db 90874 ATGTCCTGGGTGCTCTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCTCGGG 90933
 |||||

QY 61 CTGGAGGACCGTCCCGGCTCGGAGTGGTACCTGGG 98
 |||||
 Db 90934 CTGGAGGACCGTCCCGGCTCGGAGTGGTACCTGGG 90971

RESULT 33
 AY120722 1765 bp mRNA linear PLN 22-JUN-2002
 Arabidopsis thaliana putative protein (At4g35140) mRNA, complete cds.

LOCUS AY120722
 DEFINITION
 ACCESSION AY120722
 VERSION AY120722.1 GI:21539454
 KEYWORDS FLI_CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 1765)
 Southwick, A., Karlin-Neumann, G., Nguyen, M., Tripp, M., Miranda, M.,
 Palm, C. J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M. K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S. X., Narusaka, M., Pham, P. K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R. W.
 Direct Submission
 Submitted (11-JUN-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu

TITLE
 JOURNAL
 COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Southwick, A.,
 Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C. J.,
 Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M. K.,
 Kim, C., Lin, J., Liu, S. X., Pham, P. K., Sakano, H., Shinn, P.,
 Yamada, K., Ecker, J., Theologis, A. and Davis, R. W.

Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed

equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

```

FEATURES
    source
        1..1765
            /organism="Arabidopsis thaliana"
            /db_xref="taxon:3702"
            /chromosome="4"
            /clone="RAFL05-11-F17 (R15872)"
            /note="This clone is in pBluescript
            ecotype: Columbia"
        1..1765
            /gene="At4g35140"
            /protein_id="At4g35140"
            /codon_start=1
            /product="putative protein"
            /db_xref="GI:21539455"
            /translation="MSEKARRRGSNFGNHLHPVDFWREVGGISRRNFSNRFSASEN
            LVRLDEITKKLEKHCVCNTVSFNAEGDVLISGDDRRVLDWDLGNVKSFSHSGHA
            NNFOAKWPFSDDRITVTCADGVMVRASILEGDKVETSLGLHOGRAHKLCIEPGN
            PHIFTCGDLVORFEDLTAPELFTCRSDVPRRNMDATLQNALIDPNSLFA
            VGMREYARLDIRPQCEGLNGFTRADHFCPPHLIGNEDVGLTGLAFSEQLVLS
            INDEFIYFTFGMLGSPNIPSPSPISKSPVSKSESSSPKDEHSVSLYKHKNC
            TVKGNFTGPRSEYVYSGDCGRIFIRKKGGELIRVMEADRHVNCIEPHPHIVLA
            SSGIESDKVWTSKAAERATLFTPELIRKTRPRGMRYRVSPHLLAQFLQNRSS
            SPERGESSATGRELLDILTFDQSDSDENATDDDDGNSHEDFFS"
        493 a 338 c 414 g 520 t
    BASE COUNT
    ORIGIN
        Query Match      2.9%; Score 76; DB 8; Length 1765;
        Best Local Similarity 51.0%; Pred. No. 2.8e-09;
        Matches 206; Conservative 0; Mismatches 195; Indels 3; Gaps 1;

    QY 107 AATTTATCCAAAGATTAAACTTTGAAGCAACCTTAAATGTGCATGATGGTGTGTTAATA 166
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 246 ATTTGTTCTCGCTTTAGAGATATACAGNAGCTAGAGACAAAGGGTGTGTTAACA 305
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 167 CAATCTGTTGAATGACACTGGAGATATATTTATCTGCTCAGATGACACCAATAG 226
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 306 CCGTGAGTTTCAATCGGAGGAGATGTTCTGATATCTGTTCCGATGATCGCGGTGTC 365
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 227 TANTAGTATCTTACAGCAGAAAGTTTGGACAAATTCGTCAGGCGCACCGAGCAA 286
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 366 TCTTTGGGATTGGCACTTGGGAATGTGAAGCTTCTTCCATTCAGGACACGCTAATA 425
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 287 ACATATTTAGTCAAAAGTTCTTACCTTGTGACAAATGATTAACACATGATGATCTGCTCTG 346
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 426 ACGTTTTCAGCAAGTTTATGCCITCTCTGATGATGACAGATAGTAACCTGTGCTG 485
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 347 GAGATGAGTAATATTTATPACCAAGCTTGAGCAAGATGACAGAACCAACAGACATGCC 406
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 486 CTGATGGAATGTTTCGGCGTGCAGATGATCTTTCAGGCGTGAAGTGGAGACGCTTTTC- 544
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 407 AATTACGCTCATATGGAACACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 466
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 545 --TTGGGATTCATCAGGACGAGCTCATAACTATGATGATGATGATGATGATGATGATGATGAT 602
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    QY 467 CTTTCTCTCTTCTGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 510
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    Db 603 TATTTACACATGTTGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 646
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 34
AY039964
LOCUS
DEFINITION
    Arabidopsis thaliana unknown protein (At5g10940) mRNA, complete cds.
ACCESSION
    AY039964
VERSION
    AY039964.1 GI:14532679
KEYWORDS
    FLI_CDNA.
SOURCE
    Arabidopsis thaliana.
ORGANISM
    Arabidopsis thaliana

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 2594)
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Full Length cDNA Clones

Unpublished

2 (bases 1 to 2594)

Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Dale, J.M., Gibson, H.A., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Direct Submission

Submitted (11-JUN-2001) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN

Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,

Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,

Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Dale, J.M., Gibson, H.A., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Location/Qualifiers

1..2594

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAFL05-07-O21 (R10020)"

/note="This clone is in a modified pBluescript vector

(Lambda ZAP) as a XhoI/SstI insert.

ecotype: Columbia"

1..2594

/gene="At5g10940"

1..63

/gene="At5g10940"

/misc_difference 2

/gene="At5g10940"

/note="compared to genomic sequence"

/replace="C"

64..2337

/gene="At5g10940"

/codon_start=1

/evidence-experimental

/product="unknown protein"

/protein_id="AAK64141.1"

/db_xref="GI:14532680"

/translation="MDNLSFHDGSDRLINWYNSRKLHSDHGTGHANICTKPV

ELGHOGCVNALAWNSGLLSISDRLINWYNSRKLHSDHGTGHANICTKPV

PETSDRLVYSGAGDAEVRLENTSRISRAEDDNIIPSLYQCHTRRVKVLAYEPGNP

NYVMSASEGTLRQHDRETSCTPPAGTACQECRSVLLDLRSGAKPALADPKPQTL

KSCDISATRPHLLVGGSDAFRLIDRMLPPLASSKRMPFPFCVNYFCFPHLSERG

RTNLHLTHVTSPNGEVLSSYSGEHYLMNVNNGICSTGIMQYTPGDVNLFSFNN


```

/clone="NT2RP3001274"
/cell_line="NT2"
/clone_lib="teratocarcinoma"
/clone_lib="NT2RP3"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
BASE COUNT      554 a  972 c  774 g  666 t
ORIGIN

Query Match      2.7%; Score 70.6; DB 9; Length 2966;
Best Local Similarity 58.8%; Pred. No. 8.7e-08;
Matches 143; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

QY 2143 GGCATCGCACTCCAGGACATGATAAAGAGCCAAATTTCTGGG-----TGCTAAC 2196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 668 GGCACATGCAACACACACCGGATATCAAGAGGCCAAATTTCTTGGCAGCAAGCTCAG 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2197 TTGTATATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 728 TATATCGTCAGTGGCTCTGAGGATGGCTCTTCTTCATCTGGGAAAGGAGACCAACAAC 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2257 CATTGATGCTTCTGGAAGCTGATATCATGTGTGTAAGTGCCTGCGACGACATCCGTTT 2316
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 788 CTGGTCCGTGTGCTCCAAAGGGATGAGTCCATTTGTCATGCTGCGACGCCACCCGAGC 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2317 GACCAATTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGGTCACCATTAGAA 2376
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 848 TACTGCTTCTGCGCACCATGCGATCTCTGTTGCGGGCTCTGGACCCCGACCA 907
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2377 GAG 2379
|||
DB 908 GAG 910

RESULT 37
AK023778
LOCUS      Homo sapiens cDNA FLJ13716 fis, clone PLACE2000411, highly similar
to Homo sapiens epsin 2b mRNA.
ACCESSION AK023778
VERSION    AK023778.1 GI:10435813
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE     clone:PLACE2000411.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS    Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,K., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagabari,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3267)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers
1..3267
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

AK023778
LOCUS      Homo sapiens cDNA FLJ13716 fis, clone PLACE2000411, highly similar
to Homo sapiens epsin 2b mRNA.
ACCESSION AK023778
VERSION    AK023778.1 GI:10435813
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE     clone:PLACE2000411.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS    Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,K., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagabari,K., Masuho,Y. and Oshima,A.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3267)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
Location/Qualifiers
1..3267
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

Query Match      2.7%; Score 70.6; DB 9; Length 4211;
Best Local Similarity 58.8%; Pred. No. 8.7e-08;
Matches 143; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

QY 2143 GGCATCGCACTCCAGGACATGATAAAGAGCCAAATTTCTGGG-----TGCTAAC 2196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 668 GGCACATGCAACACACCGGATATCAAGAGGCCAAATTTCTTGGCAGCAAGCTCAG 727
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2197 TTGTATATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 728 TATATCGTCAGTGGCTCTGAGGATGGCTCTTCTTCATCTGGGAAAGGAGACCAACAAC 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2257 CATTGATGCTTCTGGAAGCTGATATCATGTGTGTAAGTGCCTGCGACGACATCCGTTT 2316
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 788 CTGGTCCGTGTGCTCCAAAGGGATGAGTCCATTTGTCATGCTGCGACGCCACCCGAGC 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2317 GACCAATTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGGTCACCATTAGAA 2376
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 848 TACTGCTTCTGCGCACCATGCGATCTCTGTTGCGGGCTCTGGACCCCGACCA 907
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2377 GAG 2379
|||
DB 908 GAG 910

RESULT 38
AX370510
LOCUS      AX370510      4211 bp      DNA      linear      PAT 16-FEB-2002
DEFINITION Sequence 29 from Patent WO0196371.
ACCESSION AX370510
VERSION    AX370510.1 GI:18857552
KEYWORDS   human.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
AUTHORS    Broenner,G., Ciossek,T., Dohrmann,C., Haeder,T. and Rothe,M.
Adipose-related gene
Patent: WO 0196371-A 29 20-DEC-2001;
Develon AG (DE)
Location/Qualifiers
1..4211
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT      830 a 1331 c 1147 g  903 t
ORIGIN

Query Match      2.7%; Score 70.6; DB 6; Length 4211;
Best Local Similarity 58.8%; Pred. No. 8.9e-08;
Matches 143; Conservative 0; Mismatches 94; Indels 6; Gaps 1;

QY 2143 GGCATCGCACTCCAGGACATGATAAAGAGCCAAATTTCTGGG-----TGCTAAC 2196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1908 GCCCACTGCAACACACACCGGATATCAAGAGGCCAAATTTCTTTGGCAGCAACGCTCAG 1967
Qy 2197 TTTGTAATGAGTGGTCTGACGTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
Db 1968 TATATCGTCACTGGGCTCTGACGATGCTCTCTTCATCTGGGAAAGAGAGACACCAAC 2027
Qy 2257 CATTTGATGCTTCTGGAAGCTGATATCATGTTGGTAACTGCTGCGAGCCACATCCGTTT 2316
Db 2028 CTGGTCCGTGCTCTCAAGAGGATGAGTCCATTTGCAACTGCTGCTGAGCCACACCCAGC 2087
Qy 2317 GACCAATTTTACGCTCATCTGGCATAGATATGACATAAAGATCTGTCACCATTAGAA 2376
Db 2088 TACTGCTTCTGGCCACGACGATGCTGCTGTTGTCGGCTCTGGAACCCCGACCA 2147
Qy 2377 GAG 2379
Db 2148 GAG 2150

RESULT 39
BC032523
LOCUS BC032523 4305 bp mRNA linear PRI 26-JUN-2002
DEFINITION Homo sapiens, clone MGC:44910 IMAGE:5500384, mRNA, complete cds.
ACCESSION BC032523
VERSION BC032523.1 GI:21595707
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsougeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 69 Row: 9 Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14149658.
Location/Qualifiers
1. 4305
/organism="Homo sapiens"
/db_xref="locusid:230796"
/db_xref="taxon:9606"
/clone="MGC:44910 IMAGE:5500384"
/tissue_type="Lymph, lymphoma"
/clone_lib="NIH_MGC_85"
/lab_host="DH10B"
/feature="Vector: pCMV-SPORT6"
399. .2429
CDS

```

```

/codon_start=1
/product="Unknown (protein for MGC:44910)"
/protein_id="AAH32523.1"
/db_xref="gi:21595708"
/translation="MAKVNITRDLDLIRRIKRGALSPERRYHVTDPFIRRLGLEAEELQ
GHSGVNCLEWNEKGDILASGSDOHTIWDPLHKKLLSMHTGHTANIFSVKFLPFA
QDRILLTGADSKVHVHDLVTKETIHMGDHTNRKTRATATPMPTWASDGLIR
OYDLRENSKSEVLIDLEYCGQLVEAKLTVPDNNCLAVGASGPPVRLVDIRMIH
NHRKSMKQSPAGVHTFCDKQPLDGAAYVAGHLFKLPDYNRLRLVAVTVTF
SPNGTELLVNMGGEVYFLDTYKRPYTFELLPRKSHSSQVQNGKMTNGSVNGS
LHLSNGFRLPESRGHVSQVPELPPYLERVKQOANEAFACOOQTQAIOLYSKAVORAP
HNALXGNRAAYMKRWDGDYDALDCLKAI SLNPKHLKAFHRLARCLPELKYVAE
ALCLDDEKGFKEQAHSSACDALGRDITAFSLKNDGEEKKGGFGGAPRRLSTSRK
DSTSEDMVLRESDYQFRYCHCNTTIDKEANFFGNSAQIIVSGSDGSGFTWEK
ETNLVRLVQGDSEIVNCLQHPFSYCLATSGIDPVVRLMNPRESDELTRVVEDME
GASQANRRMNADPLEVMLLNMGYRIITGLSSGGAGASDDESSGQVQCRPS"
BASE COUNT 858 a 1343 c 1189 g 915 t
ORIGIN
Query Match 2.7% Score 70.6; DB 9; Length 4305;
Best Local Similarity 58.88; Pred. No. 9e-08;
Matches 143; Conservative 0; Mismatches 94; Indels 6; Gaps 1;
QY 2143 GGCCATCGCAACTCCAGGACAATGATAAAGAAGCCAAATTTCTGGG-----TGCTAAC 2196
Db 1989 GGCCACTGCAACACACACCGGATATCAAGAGGCCAAATTTCTTTGGCAGCAACGCTCAG 2048
QY 2197 TTTGTAATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
Db 2049 TATATCGTCACTGGCTCTGACGATGGCTCTCTTCATCTGGGAAAGAGACACCAAC 2108
QY 2257 CATTTGATGCTTCTGGAAGCTGATAATCATGTGTAACCTGCTGCAGCCACATCCGTTT 2316
Db 2109 CTGTCCTGCTGCTCCAGGGGATGATCCATTTGCTCACTGCTGCAGCCACACCCAGC 2168
QY 2317 GACCCAAATTTAGCTCATCTGCGATGATATGACATAAAGATCTGTGCTACCATTAGAA 2376
Db 2169 TACTGCTTCTGCGCCACAGTCAGTCATGCTCTGTTGGGCTCTGGAACCCCGACCA 2228
QY 2377 GAG 2379
Db 2229 GAG 2231

RESULT 40
HS295C061/c
LOCUS HS295C061 798 bp mRNA linear PRI 28-JAN-1999
DEFINITION H.sapiens gene from PAC 295C6, similar to rat P044.
ACCESSION AL035304
VERSION AL035304.1 GI:4200231
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Rhodes, S.
Direct Submission
TITLE Submitted (06-JAN-1999) E-mail contact: humquery@sanger.ac.uk
JOURNAL This sequence was generated from cDNA clones isolated using
COMMENT sequence from the bacterial clone 295C6 (297876) and EST data. The
EST sequences listed match this sequence with an identity of at
least 95% between the coordinates shown.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1/ Sanger Centre name: DJ295C6.C1.1.
Location/Qualifiers
1. .798
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q24"
join(1..169,172..191,189..408)
/feature misc_feature
/feature note="match: EST AA397692"

```


OM nucleic - nucleic search, using sw model

Run on: March 25, 2003, 20:15:03 ; Search time 593 Seconds

Title: US-09-781-693A-1
(without alignments)
Sequence: 1 atgtctgggtggtgcctca.....atgaataatgagatgagaa 2580Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2573.6	99.8	3189	22	AAF58654
2	2565.8	99.4	3208	21	AAC59068
3	2508.4	97.2	3022	22	AAF59642
4	2332.6	90.4	3420	22	AAF58655
5	2140	82.9	2448	22	AAF26955
6	1542.4	59.8	1942	24	ABK35809
7	1446.4	56.1	1952	22	ABK33744
8	1067.4	41.4	1726	22	ABK33776
9	798.4	30.9	1012	22	AAS31170

us-09-781-693a-1.rng

us-09-781-693a-1.ling

wed Mar 26 09:38:49 2003

83	60.2	2.3	773	22	AAH71465	Human cervical can
84	60	2.3	60	24	AEN33333	Human spliced tran
85	60	2.3	773	21	AAC34253	Arabidopsis thalia
86	59.4	2.3	1851	24	ABK16384	Mouse cDNA encodin
87	59.4	2.3	2034	24	ABK16367	Mouse adipose prot
88	59.4	2.3	2419	24	ABK16367	Peppermint plant o
89	55.2	2.1	626	22	ABK16210	Human prostate can
90	54.6	2.1	2679	21	ABK16210	Drosophila melanog
91	54.4	2.1	2661	23	ABL02061	Drosophila melanog
92	54.4	2.1	5024	23	ABL02060	Human cervical can
93	50.4	2.0	663	22	ABL70040	Drosophila melano
94	50.4	2.0	14392	23	ABL12692	Human cDNA clone
95	49.8	1.9	778	22	AAH06956	Human SNP oligonuc
96	49.8	1.9	3115	22	AAH14651	Nucleotide sequenc
97	49.4	1.8	22	AAH27746	DNA encoding pneum	
98	45.6	1.8	1275	20	AAZ31085	S. pneumoniae pspC
99	45.2	1.8	1464	24	AAZ31099	S. pneumoniae pspC
100	45.2	1.8	1590	20	AAZ31963	DNA encoding leuco
101	43.6	1.7	1680	20	AAZ31961	DNA encoding novel
102	42.8	1.7	1686	16	AAQ87587	Human gene signal t
103	42.6	1.7	2152	23	AAQ91411	P. patens signal t
104	42	1.6	69	16	AAZ24598	Genomic sequence e
105	41	1.6	2277	24	AAZ17965	Peppermint plant o
106	40.4	1.6	143899	24	AAH38336	Human cDNA sequenc
107	40.4	1.6	202	22	AAH87807	Human nervous syst
108	40.2	1.6	1736	22	AAH15126	Human cDNA sequenc
109	40	1.6	4685	22	ABA18291	Nucleotide sequenc
110	39.4	1.5	1295	20	AAZ31077	Yeast ADP5604-asso
111	39.4	1.5	4590	22	AAH24065	Genomic sequence o
112	39.4	1.5	2365589	24	ABA90521	Chicken leucocytos
113	39.2	1.5	3399	17	AAZ05868	Buchnera sp. genom
114	39.2	1.5	640681	24	ABA92787	CDNA of a choline
115	39	1.5	429	18	AAV05099	Human polynucleoti
116	39	1.5	722	20	AAZ33209	N-terminal truncat
117	39	1.5	1129	20	AAZ34406	DNA encoding choli
118	39	1.5	1129	20	AAZ34456	Nucleotide sequenc
119	39	1.5	1242	20	AAZ31084	Nucleotide sequenc
120	39	1.5	1242	20	AAZ31090	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAF58654
AAF58654 standard; cDNA; 3189 BP.

AC AAF58654;
XX AAF58654;
XX 27-Apr-2001 (first entry)
XX Human Repro-EN-1.0 cDNA.
XX Human Repro-EN-1.0; IBI; cytostatic; breast cancer; uterine cancer;
XX Human; Repro-EN-1.0; IBI; cytostatic; breast cancer; uterine cancer;
XX prostate cancer; epitope; ss.
XX Homo sapiens.
XX WO200107616-A1.
XX 01-FEB-2001.
XX 10-MAR-2000; 2000WO-US06742.
XX 22-JUL-1999; 99US-0359084.
XX 30-JUL-1999; 99WO-US17284.
XX 23-NOV-1999; 99US-0447399.
XX (DIAG-) DIAGNOSTIC PROD CORP.
XX El Shami AS, Menon SN, French CK;
XX WPI; 2001-182795/18.

DR	P-PSDB; AAB69556.
XX	New autoantigens Repro-EN-1.0 and IB 1 polypeptides and polynucleotides
PT	are useful for diagnosing endometriosis or as a marker for pathologic
PT	conditions such as breast, uterine or prostate cancer
XX	Claim 2; Page 22-27; 11pp; English.
PS	The present sequence is given in a specification relating to recombinant
CC	polynucleotides comprising nucleotide sequences encoding a polypeptide
CC	epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a
CC	fully defined sequence of 860 and 937 amino acids. The epitope
CC	specifically binds to antibodies from subjects diagnosed with
CC	endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for
CC	diagnosing endometriosis by detecting antibodies in immunoassays, and
CC	are used as markers for pathologic conditions e.g. breast, uterine or
CC	prostate cancer. Methods for detecting Repro-EN-1.0 or IB 1
CC	polynucleotides of polypeptides are useful in the diagnosis of these
CC	cancers, monitoring their progress or treatment, and determining patient
CC	prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1
CC	may be used as probes for detecting mRNA from cell types suspected to
CC	be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and
CC	IB 1 polypeptides and immunogenic fragments may be used as specifically
CC	controls in diagnostic assays to detect antibodies that specifically
CC	bind the proteins from patient serum samples. The polypeptides are
CC	useful as immunogens for eliciting antibody production against epitopes
CC	of the protein, and as controls in diagnostic methods.
XX	Sequence 3189 BP; 977 A; 661 C; 724 G; 827 T; 0 other;
XX	Query Match 99.8%; Score 2573.6; DB 22; Length 3189;
XX	Best Local Similarity 99.8%; Pred. No. 0;
XX	Matches 2576; Conservative 4; Mismatches 0; Gaps 0;
QY	1 ATGCTCGGGGTGCTCTTACCCACACCTTGTCTGGACGTGAGAAAGTCCCTCGGG 60
DB	176 ATGCTCGGGGTGCTCTTACCCACACCTTGTCTGGACGTGAGAAAGTCTCTCGGG 235
QY	61 CTGAGGACCCCTCCCGCTCGGAGTCTGCTGCTGGAGAGAGCAATTTATCAAGA 120
QY	236 CTGAGGACCCCTCCCGCTCGGAGTCTGCTGCTGGAGAGAGCAATTTATCAAGA 295
DB	121 TTAACCTTGAGCAACCTTAAATGTGATGATGTTGTGTATTAACAATCTGTTCGAT 180
QY	296 TTAACCTTGAGCAACCTTAAATGTGATGATGTTGTGTATTAACAATCTGTTCGAT 355
DB	181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATTAGTAATTAATCTCT 240
QY	356 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATTAGTAATTAATCTCT 415
DB	241 TACAGCAGAAAGGTTTGTGACACCAATTCGTTACGGCCAGCAGCAACATATTAGTGA 300
QY	416 TACAGCAGAAAGGTTTGTGACACCAATTCGTTACGGCCAGCAGCAACATATTAGTGA 475
DB	301 AGTTCTTACCTTGACAAATGATTAACAGATTTGATCTCTGCTGGAGATGAGTAATA 360
QY	476 AGTTCTTACCTTGACAAATGATTAACAGATTTGATCTCTGCTGGAGATGAGTAATA 535
DB	361 TTTATACCAACCTTGAGCAAGATGACAAACCAACCAACCAATTCGCTGCTGCTCTCT 420
QY	536 TTTATACCAACCTTGAGCAAGATGACAAACCAACCAACCAATTCGCTGCTGCTCTCT 595
DB	421 TATGAACTACTATGAGATTTATGACCTGTACCAATGACCTTACACTTTCTCTCTGT 480
QY	596 TATGAACTACTATGAGATTTATGACCTGTACCAATGACCTTACACTTTCTCTCTGT 655
DB	481 GGTGAGATGGAACCTTTAGTGGTTTGTATACGCAATCAAACTAGCTGCAACAAAGAA 715
QY	656 GGTGAGATGGAACCTTTAGTGGTTTGTATACGCAATCAAACTAGCTGCAACAAAGAA 800
DB	541 GATTGTAAGATGATATTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
DB	716 GATTGTAAGATGATATTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT

QY 601 CCACCAATACCATATTAACCTTCTGCTGTTGTTCTGACAGCTCAGTACCAATATATGAT 660
 Db 776 CCACCAATACCATATTAACCTTCTGCTGTTGTTCTGACAGCTCAGTACCAATATATGAT 835
 QY 661 CGGCGAATGCTGGGCACAGAGCTACAGGGAATATATGAGTCTGAGGAGTCTGGAATG 720
 Db 836 CGGCGAATGCTGGGCACAGAGCTACAGGGAATATATGAGTCTGAGGAGTCTGGAATG 895
 QY 721 GTTCCCGCTTTATTCCTCCCATCTTAATAAAGCTCTGAGAGTACATCTCTGCT 780
 Db 896 GTTCCCGCTTTATTCCTCCCATCTTAATAAAGCTCTGAGAGTACATCTCTGCT 955
 QY 781 TACAGTGAAGATGCTCAAGAGATCTGTTAGTTACTCTTCAGATATACATATCTTTT 840
 Db 956 TACAGTGAAGATGCTCAAGAGATCTGTTAGTTACTCTTCAGATATACATATCTTTT 1015
 QY 841 GACCCGAAGATGATACAGCAGAGACTTAAACTCTCTTCTGCGGAGAGAGAGAA 900
 Db 1016 GACCCGAAGATGATACAGCAGAGACTTAAACTCTCTTCTGCGGAGAGAGAGAA 1075
 QY 901 GAGTTGCGAACAACCACTTAAGCTTTGAGACTTCGTGTTGATTTGTTGATCTGTTG 960
 Db 1076 GAGTTGCGAACAACCACTTAAGCTTTGAGACTTCGTGTTGATTTGTTGATCTGTTG 1135
 QY 961 CCCAGAGCAAGCGGAGAGTGAACGAGAGATGAGAGCAGAGTCCCAATCTGGA 1135
 Db 1136 CCCAGAGCAAGCGGAGAGTGAACGAGAGATGAGAGCAGAGTCCCAATCTGGA 1200
 QY 1021 TTGATGAGAGATGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTCAAGTGA 1195
 Db 1196 TTGATGAGAGATGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTCAAGTGA 1255
 QY 1081 CAAAGCAATAGAGAGAGAGATGCTGAGCAGAGTGGAAAGTCAATCAATAT 1140
 Db 1256 CAAAGCAATAGAGAGAGATGCTGAGCAGAGTGGAAAGTCAATCAATAT 1315
 QY 1141 TCAACTCTTCTTACGCTCCATCAAGTCTGATTTGAGAGTGAAGTCAATCAAT 1200
 Db 1316 TCAACTCTTCTTACGCTCCATCAAGTCTGATTTGAGAGTGAAGTCAATCAAT 1375
 QY 1201 GTAGATCTCCAGCTGAACAAATTTCTCAGCTTCTACATCTCTTCAATGTCAGT 1260
 Db 1376 GTAGATCTCCAGCTGAACAAATTTCTCAGCTTCTACATCTCTTCAATGTCAGT 1435
 QY 1261 GCTCATTCGACATCTCTCCACAGAGCCCTCATCTACTCTTCTGATCTCTTCTCA 1320
 Db 1436 GCTCATTCGACATCTCTCCACAGAGCCCTCATCTACTCTTCTGATCTCTTCTCA 1495
 QY 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCTGAT 1380
 Db 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCTGAT 1555
 QY 1381 AACAAATGAAGAGTGAAGCCCAACAGGAGAGTGAACAGTGAACAGTGAACAG 1440
 Db 1556 AACAAATGAAGAGTGAAGCCCAACAGGAGAGTGAACAGTGAACAGTGAACAG 1500
 QY 1441 TACAGCAGAGAGAGAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAG 1615
 Db 1616 TACAGCAGAGAGAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAG 1675
 QY 1501 AGTATAGCATCAAGTCTGAGAGATTTGGAGCCATTTGCAATCTGAGGCTCAGGAG 1560
 Db 1676 AGTATAGCATCAAGTCTGAGAGATTTGGAGCCATTTGCAATCTGAGGCTCAGGAG 1735
 QY 1561 TCTTCTGCTCCACAGCTCAGTGAACAGGAGAGTGAACAGTGAACAGTGAACAG 1620
 Db 1736 TCTTCTGCTCCACAGCTCAGTGAACAGGAGAGTGAACAGTGAACAGTGAACAG 1795
 QY 1621 GAAGATCATCAGAGATGAGCAAAATATCAGAGAGAGTATCTGCAAGAAACCACT 1680
 Db 1796 GAAGATCATCAGAGATGAGCAAAATATCAGAGAGAGTATCTGCAAGAAACCACT 1855

QY 1681 GAGAACCATATCAATATATACCAATACAGATACAGTTCACAGCCAAAGCCATTCGATTC 1740
 Db 1856 GAGAACCATATCAATATATACCAATACAGATACAGTTCACAGCCAAAGCCATTCGATTC 1915
 QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTGCTGCTTGTGGGTTCCAGAGAAATCTGCT 1800
 Db 1916 TCAGGAGAAAGAAATGACCTCAATCTTGTGCTGCTTGTGGGTTCCAGAGAAATCTGCT 1975
 QY 1801 TCATCTCAAAAGAGCCAAAGAACCAAGAACTTCAGATCAGCTAGCACTGAGAGTCTGCT 1860
 Db 1976 TCATCTCAAAAGAGCCAAAGAACCAAGAACTTCAGATCAGCTAGCACTGAGAGTCTGCT 2035
 QY 1861 AATGAAATTAACCACTCACTGAGCTCAGTTCCTCAACAGAAAGCCACTGGCCCTTCAGCT 1920
 Db 2036 AATGAAATTAACCACTCACTGAGCTCAGTTCCTCAACAGAAAGCCACTGGCCCTTCAGCT 2095
 QY 1921 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGATGAC 1980
 Db 2096 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGATGAC 2155
 QY 1981 CCAGTCTGTATCCCAAGTTCAGAGTATCGAGCAGGAGCTGCTGATAGACGCTCTGCTGT 2040
 Db 2156 CCAGTCTGTATCCCAAGTTCAGAGTATCGAGCAGGAGCTGCTGATAGACGCTCTGCTGT 2215
 QY 2041 GCCCGTATTCAGAGTTCCTTCAGACGAGAAAGAAAGGAAAGAAATGAGAAATGAGAT 2100
 Db 2216 GCCCGTATTCAGAGTTCCTTCAGACGAGAAAGAAAGGAAAGAAATGAGAAATGAGAT 2275
 QY 2101 ACTTTGAACATTAAGAGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTCCAGG 2160
 Db 2276 ACTTTGAACATTAAGAGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTCCAGG 2335
 QY 2161 ACAATGATTAAGAGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTCCAGG 2220
 Db 2336 ACAATGATTAAGAGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTCCAGG 2395
 QY 2221 GCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2380
 Db 2396 GCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2455
 QY 2281 AATCATGTTGTTAACTGCTGACGACCATCTGTTGACCCAAATTTAGCCTCATCTGCTG 2340
 Db 2456 AATCATGTTGTTAACTGCTGACGACCATCTGTTGACCCAAATTTAGCCTCATCTGCTG 2515
 QY 2341 ATAGATTATGACATTAAGATCTGCTACCATTAAGAGAGTCAAGGATTTTAAACGAA 2400
 Db 2516 ATAGATTATGACATTAAGATCTGCTACCATTAAGAGAGTCAAGGATTTTAAACGAA 2575
 QY 2401 CTTGCTGATGAATTAATACTCGAAACGACATCTGCTGGAAGAACTAGAAACACCAAT 2460
 Db 2576 CTTGCTGATGAATTAATACTCGAAACGACATCTGCTGGAAGAACTAGAAACACCAAT 2635
 QY 2461 ACAGTTCCAGCTCTTTCATGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGAC 2520
 Db 2636 ACAGTTCCAGCTCTTTCATGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGAC 2695
 QY 2521 CGSTTGGAGGTGACAGATCAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 2580
 Db 2696 CGSTTGGAGGTGACAGATCAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 2755

RESULT 2

AAC59068

ID AAC59068 standard; cdna; 3208 BP.

XX AAC59068;

XX 02-FEB-2001 (first entry)

XX Human secreted protein coding sequence SEQ ID NO: 30.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;

KW

us-09-781-693a-1.rng

Wed Mar 26 09:38:49 2003

KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurologic disease; infection; human; secreted protein; ss.
XX Homo sapiens.
OS WO200055171-A1.
PN 21-SEP-2000.
PD 09-MAR-2000; 2000WO-US06043.
XX 12-MAR-1999; 99US-0124146.
XX 23-NOV-1999; 99US-0167061.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-638174/61.
XX P-PSDB; AAB27926.
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX Claim 1; Page 344-345; 438pp; English.

XX The invention relates to the isolation of genes AA59049-A59098 encoding
XX 50 human secreted proteins AAB27907-A27956. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion (SBO1D1) for increasing the stability of
XX the fusion protein as compared to the human protein only. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated
XX from a range of human tissues disclosed in the specification. The
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
XX wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.

XX Sequence 3208 BP; 975 A; 667 C; 739 G; 827 T; 0 other;
XX

Query Match 99.4%; Score 2565.8; DB 21; Length 3208;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2578; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 ATGCTCGGGTGGCTCTACACACACCTGTGTGGGACGTGAGAAAGGTCCTCGGG 60
DB 207 ATGCTCGGGTGGCTCTACACACACCTGTGTGGGACGTGAGAAAGGTCCTCGGG 266
QY 61 CTGGAGGACCGTCCGGCTCGGAGTCGCTACCTGGGAAGAGAGAAATTTATCCAAAGA 120
DB 267 CTGGAGGACCGTCCGGCTCGGAGTCGCTACCTGGGAAGAGAGAAATTTATCCAAAGA 326
QY 121 TTAACAACCTTGAAGCAACCTTTAATGTCATGATGTTGTGTTAATCAATCTGTTGGAAT 180
DB 327 TTAACAACCTTGAAGCAACCTTTAATGTCATGATGTTGTGTTAATCAATCTGTTGGAAT 386
QY 181 GACACTGGAGATATATATTTATCTGGCTCAGATGACCAATTTAGTAATTTAGTAATCTCT 240
DB 387 GACACTGGAGATATATATTTATCTGGCTCAGATGACCAATTTAGTAATTTAGTAATCTCT 446
QY 241 TACAGCAAGAAAGTTTGGACAAATTCCTTCAGGGCCACCGAGCAACATATTTAGTGCA 300
DB 447 TACAGCAAGAAAGTTTGGACAAATTCCTTCAGGGCCACCGAGCAACATATTTAGTGCA 506

QY 301 AAGTCTTACCTTGTACAAATGATTAACAGATGTTATCCTGCTCTGGAGATGGAGTAATA 360
DB 507 AAGTCTTACCTTGTACAAATGATTAACAGATGTTATCCTGCTCTGGAGATGGAGTAATA 566
QY 361 TTTTATACCAACGTTGAGCAAGATGCAAGAAACCAACAGACAAATTTACGTTGCA 419
DB 567 TTTTATACCAACGTTGAGCAAGATGCAAGAAACCAACAGACAAATTTACGTTGCA 626
QY 420 TTTATGGAATCTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 479
DB 627 TTTATGGAATCTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 686
QY 480 TGGTGAAGATGGAATCTTATGAGTGTGTTGATACAGGATCAAACTAGCTGCACAAAAGA 539
DB 687 TGGTGAAGATGGAATCTTATGAGTGTGTTGATACAGGATCAAACTAGCTGCACAAAAGA 746
QY 540 AGATGTAAAGATGATATTTAATTAACCTGTCAGCTGTCGAGCTGTCGAGTGTGTTGTTG 599
DB 747 AGATGTAAAGATGATATTTAATTAACCTGTCAGCTGTCGAGCTGTCGAGTGTGTTGTTG 806
QY 600 CCCACCAATACCATATTTACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 859
DB 807 CCCACCAATACCATATTTACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 866
QY 660 TCGGGGATGCTGGGCAACAGAGCTACAGGGAATTTATGACAGCTCAGTACGATATATGA 719
DB 867 TCGGGGATGCTGGGCAACAGAGCTACAGGGAATTTATGACAGCTCAGTACGATATATGA 779
QY 720 GGTTCGCCGTTTATTCCTTCCCATCTTAATAAGTCTCTGACAGTGCATCTCTGTTG 798
DB 927 GGTTCGCCGTTTATTCCTTCCCATCTTAATAAGTCTCTGACAGTGCATCTCTGTTG 839
QY 780 TTAGCTGAAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1046
DB 987 TTAGCTGAAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 899
QY 840 TGACCCGAAAGATGATACAGGAGCTACAGGAGCTTAAATCTCTTCGCGGAGAGAGAGA 1106
DB 1047 TGACCCGAAAGATGATACAGGAGCTACAGGAGCTTAAATCTCTTCGCGGAGAGAGAGA 959
QY 900 AGATGTGCGCAACACCACTTAAGCTTTGACATTCCTGTTGTTGTTGTTGTTGTTGTTGTTG 1166
DB 1107 AGATGTGCGCAACACCACTTAAGCTTTGACATTCCTGTTGTTGTTGTTGTTGTTGTTGTTG 1019
QY 960 ACCGAGCAAGGCGGAGAGTGAACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1226
DB 1167 ACCGAGCAAGGCGGAGAGTGAACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1079
QY 1020 ATTGATGAGAGATGCTGATATGTTTATCAAGATGTTTGAAGAGAGAGAGAGAGAGAGAG 1286
DB 1227 ATTGATGAGAGATGCTGATATGTTTATCAAGATGTTTGAAGAGAGAGAGAGAGAGAGAG 1139
QY 1080 ACAAAGCAATAGAT 1346
DB 1287 ACAAAGCAATAGAT 1199
QY 1140 TTCAACTCTTCTACGTCCTCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGA 1406
DB 1347 TTCAACTCTTCTACGTCCTCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGA 1259
QY 1200 AGTATGATCTCCAGCTGAACAATTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1466
DB 1407 AGTATGATCTCCAGCTGAACAATTTCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1319
QY 1260 GGCTCATCTGACATCATCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1526
DB 1467 GGCTCATCTGACATCATCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1379
QY 1320 AGACAGTGAACAAAGGAGTCTGTTGAGGATCTGAGGATCTGAGGATCTGAGGATCTGAGG 1586
DB 1527 AGACAGTGAACAAAGGAGTCTGTTGAGGATCTGAGGATCTGAGGATCTGAGGATCTGAGG 1439
QY 1380 TACAAATATGAAGAGTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTTAA 1439

Db 1587 TAACAAATTAATGAAGAGCTTGAGCCCAACACAGGACAGGTGAACAGTTTAAAGTTTCA 1646
QY 1440 CTACAGCACAGAGGAACAATACAAAGCACAAATAAACTGAACCTTACAGATGAATGAG 1499
Db 1647 CTACAGCACAGAGGAACAATACAAAGCACAAATAAACTGAACCTTACAGATGAATGAG 1706
QY 1500 CAGTATAGCATCAAGTTCTAGAGGAATGGAGCCATTGCAAACTCTGAGGTCAGGAGGA 1559
Db 1707 CAGTATAGCATCAAGTTCTAGAGGAATGGAGCCATTGCAAACTCTGAGGTCAGGAGGA 1766
QY 1560 ATCTTTCTGCCACAGAGCTCAGTGCACACCAAGAGGACAGTGAACAAAGCTCC 1619
Db 1767 ATCTTTCTGCCACAGAGCTCAGTGCACACCAAGAGGACAGTGAACAAAGCTCC 1826
QY 1620 TGAAGATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAACCCAGT 1679
Db 1827 TGAAGATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAACCCAGT 1886
QY 1680 TGAGAACCATATCAATATACCAATACAGATAGTTTCACAGCCAGCCATTGGATTCCAA 1739
Db 1887 TGAGAACCATATCAATATACCAATACAGATAGTTTCACAGCCAGCCATTGGATTCCAA 1946
QY 1740 CTCAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCACAGAAATCTGC 1799
Db 1947 CTCAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCACAGAAATCTGC 2006
QY 1800 TTCATCTGAAAAGCCAGAGCAAGAACTTCAGATCAGATAGCAGTGAAGTGTCTAC 1859
Db 2007 TTCATCTGAAAAGCCAGAGCAAGAACTTCAGATCAGATAGCAGTGAAGTGTCTAC 2066
QY 1860 CANTGAAATACCAATCTGAGCCCTCAGTTCACCAACAGAGCCACTGGGCCCTCAGC 1919
Db 2067 CANTGAAATACCAATCTGAGCCCTCAGTTCACCAACAGAGCCACTGGGCCCTCAGC 2126
QY 1920 TCATGAGAAATACCAATCTGAGCCCTCAGTTCACCAACAGAGCCACTGGGCCCTCAGC 1979
Db 2127 TCATGAGAAATACCAATCTGAGCCCTCAGTTCACCAACAGAGCCACTGGGCCCTCAGC 2186
QY 1980 CCGAGTCTGATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGTATAGCCCTCTGCTGT 2039
Db 2187 CCGAGTCTGATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGTATAGCCCTCTGCTGT 2246
QY 2040 TGCCGCTATTTCAGGAGTCTTCAGAGCGGAGAAAGAAAGAAATGGAAGAAATTTGGA 2099
Db 2247 TGCCGCTATTTCAGGAGTCTTCAGAGCGGAGAAAGAAAGAAATGGAAGAAATTTGGA 2306
QY 2100 TACTTTGAACATTTAGAGGCGCTAGTAAATAATGGTTTATAAGGCGCATCGCAACTCCAG 2159
Db 2307 TACTTTGAACATTTAGAGGCGCTAGTAAATAATGGTTTATAAGGCGCATCGCAACTCCAG 2366
QY 2160 GACAATGATAAAGAACCCATTTCTGGGTGCTAACTTTGTATAGTGGTTCTGACTG 2219
Db 2367 GACAATGATAAAGAACCCATTTCTGGGTGCTAACTTTGTATAGTGGTTCTGACTG 2426
QY 2220 TGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGAAAGCTGA 2279
Db 2427 TGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGAAAGCTGA 2486
QY 2280 TAATCATGTTGTAAGTGGTCTGAGCCACATCCGTTTGTACCCAAATTTAGCTCTCATCTGG 2339
Db 2487 TAATCATGTTGTAAGTGGTCTGAGCCACATCCGTTTGTACCCAAATTTAGCTCTCATCTGG 2546
QY 2340 CATGATTATGACATAAAGATCTGCTGACCATTTAGAGGATCAAGGATTTTAAACCGAAA 2399
Db 2547 CATGATTATGACATAAAGATCTGCTGACCATTTAGAGGATCAAGGATTTTAAACCGAAA 2606
QY 2400 ACTTCTCATGAGTTTAACTCGAAGAACTCATGCTGGAAGAACTAGAAACACCAT 2459
Db 2607 ACTTCTCATGAGTTTAACTCGAAGAACTCATGCTGGAAGAACTAGAAACACCAT 2666
QY 2460 TACAGTTCAGCCCTCTTTCATGTTGAGGATGTTGGCTTCACCTTATCATATCCGAGCTGA 2519

Db 2667 TACAGTTCAGCCCTCTTTCATGTTGAGGATGTTGGCTTCACCTTATCATATCCGAGCTGA 2726
QY 2520 CCGGTTGAGGGTGACAGATCAGAAGGCTCTGGTCAAGAGAAATGAAATGAGATGAGGA 2579
Db 2727 CCGGTTGAGGGTGACAGATCAGAAGGCTCTGGTCAAGAGAAATGAAATGAGATGAGGA 2786
QY 2580 A 2580
Db 2787 A 2787

RESULT 3
AAF59642
ID AAF59642 standard; cDNA; 3022 BP.
XX
AC AAF59642;
XX
DT 24-APR-2001 (first entry)
XX
DE Human cell cycle and proliferation protein CCYPR-53 cDNA, SEQ ID NO:107.
XX
KW Cell cycle and proliferation protein; CCYPR; human; agonist;
KW antagonist; gene therapy; detection; gene therapy;
KW transgenic animal disease model; immune disorder;
KW developmental disorder; cell signalling disorder;
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW arteriosclerosis; asthma; allergy; diabetes mellitus;
KW menstrual cycle disorder; bacterial infection; ss.
XX
OS Homo sapiens.
XX
PN WO200107471-A2.
XX
PD 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US19948.
XX
PF 21-JUL-1999; 99US-0145075.
XX
PR 08-SEP-1999; 99US-0153129.
PR 10-NOV-1999; 99US-0164647.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAN, Baughn MR, Patterson C, Shah P;
XX
DR WPI: 2001-112727/12.
XX
PT P-PSDB; AAB60505.
XX
PT Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signalling disorders and cell proliferative disorders including cancer -
PS
PS Claim 5; Page 204; 205pp; English.
XX
CC Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented with CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative

CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.

XX	Sequence	3022 BP;	937 A;	611 C;	674 G;	800 T;	0 other;	
SO	Query Match	97.2%;	Score	2508.4;	DB	22;	Length	3022;
	Best Local Similarity	97.7%;	Pred. No.	0;				
	Matches	2579;	Conservative	1;	Indels	60;	Gaps	1;
QY	1	ATGCTCTCGGGTGGCTCCCTACCCACACACCTGTTCTGGGACGCTGAGGAAAGTCCCTCGGG	60					
DB	119	ATGCTCTCGGGTGGCTCCCTACCCACACCTGTTCTGGGACGCTGAGGAAAGTCCCTCGGG	178					
QY	61	CTGGAGACCCGCTCCCGGCTCGGGATCGCTGCTGGGAGAGAGAAATTTATCCAAAGA	120					
DB	179	CTGGAGACCCGCTCCCGGCTCGGGATCGCTGCTGGGAGAGAGAAATTTATCCAAAGA	238					
QY	121	TTAAACTTCAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAATCTGTGGAAAT	180					
DB	239	TTAAACTTCAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAATCTGTGGAAAT	298					
QY	181	GACACTGGAGATATATTTTATCTGGCTAGATGACACCAAAATAGTAAATAGTAAATCT	240					
DB	299	GACACTGGAGATATATTTTATCTGGCTAGATGACACCAAAATAGTAAATAGTAAATCT	358					
QY	241	TACAGCAGAAAGGTTTGAACAATTCGTTTCAGGGCAGCCAGCAACATATTTAGTGCA	300					
DB	359	TACAGCAGAAAGGTTTGAACAATTCGTTTCAGGGCAGCCAGCAACATATTTAGTGCA	418					
QY	301	AAGTCTTACCTTGTACAAATGATAACAGATGTTATCTCTGCTGAGATGAGTAATA	360					
DB	419	AAGTCTTACCTTGTACAAATGATAACAGATGTTATCTCTGCTGAGATGAGTAATA	478					
QY	361	TTTATACCAAGTGTGACAGATGACAGAAACCAACAGACAAATGCCAATTTACGTGTCAT	420					
DB	479	TTTATACCAAGTGTGACAGATGACAGAAACCAACAGACAAATGCCAATTTACGTGTCAT	538					
QY	421	TATGGAATCTACTTATGAGATATGACTGTACCAATGACCTTACACTTTCTCTGT	480					
DB	539	TATGGAATCTACTTATGAGATATGACTGTACCAATGACCTTACACTTTCTCTGT	598					
QY	481	GGTGAAGTGAACCTGTAGTGGTGTGATACACGCATCAAAACTAGCTGCACAAAGAA	540					
DB	599	GGTGAAGTGAACCTGTAGTGGTGTGATACACGCATCAAAACTAGCTGCACAAAGAA	658					
QY	541	GATTTGTAAGATGATATTTTAAATTAAGTGTGCTGACAGCTCAGTACGAATATATGAT	600					
DB	659	GATTTGTAAGATGATATTTTAAATTAAGTGTGCTGACAGCTCAGTACGAATATATGAT	718					
QY	601	CCACCAATACCATATTTACCTTGTGTTGTTCTGACAGCTCAGTACGAATATATGAT	660					
DB	719	CCACCAATACCATATTTACCTTGTGTTGTTCTGACAGCTCAGTACGAATATATGAT	778					
QY	661	CGGCGAATGCTGGGCACAGAGAGTACAGGAATATGCGAGTTCGAGGACTACTGGAATG	720					
DB	779	CGGCGAATGCTGGGCACAGAGAGTACAGGAATATGCGAGTTCGAGGACTACTGGAATG	838					
QY	721	GTTCCCGGTTTATCTCCCTCCATCTTAATTAAGTCTCTGACAGTGCATCTCTGT	780					
DB	839	GTTCCCGGTTTATCTCCCTCCATCTTAATTAAGTCTCTGACAGTGCATCTCTGT	898					
QY	781	TACAGTGAAGATGCTCAAGAGATTCCTCGTTAGTACTCTTCAGATACATATCTTTT	840					
DB	899	TACAGTGAAGATGCTCAAGAGATTCCTCGTTAGTACTCTTCAGATACATATCTTTT	958					
QY	841	GACCCGAAGATGATACAGCAGAGAACTTAAACTCTCTGCGGAGAGAGAGAGAA	900					
DB	959	GACCCGAAGATGATACAGCAGAGAACTTAAACTCTCTGCGGAGAGAGAGAGAA	1018					
QY	901	GAGTTGCGACAAACCACTAGTAAAGGTTGAGACTTGGTGGATCTGATCTGGA	960					

DB	1019	GAGTTGCGACAAACCCAGCTTAAGCGTTTGAAGATTCGTTGATTTGTCAGACTATGGA	1078					
QY	961	CCAGAGCAAGCGCGGAGAGTGAACAGAACAGAGTGAAGAGAGTCCCAATGTGCA	1020					
DB	1079	CCAGAGCAAGCGCGGAGAGTGAACAGAACAGAGTGAAGAGAGTCCCAATGTGCA	1138					
QY	1021	TTGATGCGAGAAATGCTGATATGTTTATCAAGATGTTTGAAGAAGCAAGTGAAGTTGA	1080					
DB	1139	TTGATGCGAGAAATGCTGATATGTTTATCAAGATGTTTGAAGAAGCAAGTGAAGTTGA	1198					
QY	1081	CAAAAGCAATAGAGAGAGAGAGTCTGACCCAGAGTGAAGTCAATCAGATAT	1140					
DB	1199	CAAAAGCAATAGAGAGAGAGAGTCTGACCCAGAGTGAAGTCAATCAGATAT	1258					
QY	1141	TCAACTCTTCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTCAATGGA	1200					
DB	1259	TCAACTCTTCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTCAATGGA	1318					
QY	1201	GTAGATACCTCCAGGTGAACAAATTTCTCAGCCTTCTACATCTCTACATCTCTACAT	1260					
DB	1319	GTAGATACCTCCAGGTGAACAAATTTCTCAGCCTTCTACATCTCTACATCTCTACAT	1378					
QY	1261	GCTCATTCGACATATCTCCACAGAAAGCCCTCATTTCTCTCTCTCTCTCTCTCTCTCA	1320					
DB	1379	GCTCATTCGACATATCTCCACAGAAAGCCCTCATTTCTCTCTCTCTCTCTCTCTCTCA	1438					
QY	1321	GACAGTGAACAAAGGAGTCTGTTGAGGATCTGACACCAACACATCATCAGTCT---	1377					
DB	1439	GACAGTGAACAAAGGAGTCTGTTGAGGATCTGACACCAACACATCATCAGTCTGAT	1498					
QY	1378	-----GAT	1380					
DB	1499	TCCT	1558					
QY	1381	AACAATTAATGAAAGTGAAGCCCAACAGGACAGTGAACAGTCTTAAAGTTGAC	1440					
DB	1559	AACAATTAATGAAAGTGAAGCCCAACAGGACAGTGAACAGTCTTAAAGTTGAC	1618					
QY	1441	TACAGCAGAAAGGAACTACAAGCACAATAAACTGAACCTTTACAGATGAATGGAGC	1500					
DB	1619	TACAGCAGAAAGGAACTACAAGCACAATAAACTGAACCTTTACAGATGAATGGAGC	1678					
QY	1501	AGTATAGCTCAAGTCTAGAGGAATTTGAGGCAATGCAAACTCTGAGGTCAGAGAGAA	1560					
DB	1679	AGTATAGCTCAAGTCTAGAGGAATTTGAGGCAATGCAAACTCTGAGGTCAGAGAGAA	1738					
QY	1561	TCCTTCCTCCACAGAGTCTGAGTGAACACCAAGAGAGACAGTGAACAAAGTCTCT	1620					
DB	1739	TCCTTCCTCCACAGAGTCTGAGTGAACACCAAGAGAGACAGTGAACAAAGTCTCT	1798					
QY	1621	GAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGAGTATCTGAGAAAACCCAGTT	1680					
DB	1799	GAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGAGTATCTGAGAAAACCCAGTT	1858					
QY	1681	GAGAACCATATCAATATAACACAATCAGATAAGTTTCAGAGCAAGCCATTTGGATTTCAAC	1740					
DB	1859	GAGAACCATATCAATATAACACAATCAGATAAGTTTCAGAGCAAGCCATTTGGATTTCAAC	1918					
QY	1741	TCAGGAGAAAGAAATGACCTCAATCTGCTCTTCTGGGTTCCAGAGAAATCTGCT	1800					
DB	1919	TCAGGAGAAAGAAATGACCTCAATCTGCTCTTCTGGGTTCCAGAGAAATCTGCT	1978					
QY	1801	TCATCTGAAGAAAGCCAGGAACCCAGAACTTCAGATCAGACTAGACTAGAGTGTCTACC	1860					
DB	1979	TCATCTGAAGAAAGCCAGGAACCCAGAACTTCAGATCAGACTAGACTAGAGTGTCTACC	2038					
QY	1861	AATGAAATTAACACCAATCTGAGGCTCAGTTCGAAACAGAGCCAGTGGGCTTCAGCT	1920					
DB	2039	AATGAAATTAACACCAATCTGAGGCTCAGTTCGAAACAGAGCCAGTGGGCTTCAGCT	2098					
QY	1921	CATGAGAAACATCTCCAGGAGTCTGCTCTCTCAGGACACAGATGACAGTGTGATGAC	1980					
DB	2099	CATGAGAAACATCTCCAGGAGTCTGCTCTCTCAGGACACAGATGACAGTGTGATGAC	2158					

QY	1981	CCAGTCTCTGATCCACAGTGC	AAAGTATCTC	GAGCAGGACCTCGT	GTATGACGCTCTGCTGTT	2040
Db	2159	CCAGTCTCTGATCCACAGTGC	AAAGTATCTC	GAGCAGGACCTCGT	GTATGACGCTCTGCTGTT	2218
QY	2041	GCCCGTATTCAGGAGTTC	TTCAGACGGAGAAA	AGAAAGAAAGAAATGGA	AGAAATTTGGAT	2100
Db	2219	GCCCGTATTCAGGAGTTC	TTCAGACGGAGAAA	AGAAAGAAAGAAATGGA	AGAAATTTGGAT	2278
QY	2101	ACTTTGAAACATTAGAAGCG	CGCTAGTAAAAATGGTTAT	ATAAAGGCCATCGCAACTCCAGG		2160
Db	2279	ACTTTGAAACATTAGAAGCG	CGCTAGTAAAAATGGTTAT	ATAAAGGCCATCGCAACTCCAGG		2338
QY	2161	ACAATGATAAAAAGACCCRA	TTTCTCGGGTGCTAAC	TTTGTAAATGAGTGGTTCTGACTGT		2220
Db	2339	ACAATGATAAAAAGACCCRA	TTTCTCGGGTGCTAAC	TTTGTAAATGAGTGGTTCTGACTGT		2398
QY	2221	GGCCACATTTTCATCTGGG	ATCGGCACACTGCTGAGCA	TTTTCATCTCTCGAAGCTCAT		2280
Db	2399	GGCCACATTTTCATCTGGG	ATCGGCACACTGCTGAGCA	TTTTCATCTCTCGAAGCTCAT		2458
QY	2281	AATCATGTGGTAAACTGCTG	CAGCCACATCCGTTTGACC	CAATTTTACCCCTCATCTGGC		2340
Db	2459	AATCATGTGGTAAACTGCTG	CAGCCACATCCGTTTGACC	CAATTTTACCCCTCATCTGGC		2518
QY	2341	ATAGATTATGACATAAAGAT	CTGGTCACCATTAGAAGAGT	CAAGGATTTTAAACCGA		2400
Db	2519	ATAGATTATGACATAAAGAT	CTGGTCACCATTAGAAGAGT	CAAGGATTTTAAACCGA		2578
QY	2401	CTTGCTGATGAAGTTATAC	TCTGGAACCACTCATGCTG	GGAAGAACTAGAAACACCAT		2460
Db	2579	CTTGCTGATGAAGTTATAC	TCTGGAACCACTCATGCTG	GGAAGAACTAGAAACACCAT		2638
QY	2461	ACAGTTCCAGGCTCTTTCAT	GTGAGGATGTGGCTTCACT	TATATCATATCCGAGCTGAC		2520
Db	2639	ACAGTTCCAGGCTCTTTCAT	GTGAGGATGTGGCTTCACT	TATATCATATCCGAGCTGAC		2598
QY	2521	CGGTTGAGGGTGACAGATCA	GAGAGCTCTGGTCAAGAGAT	GAAATGAGGTATGAGGAA		2580
Db	2699	CGGTTGAGGGTGACAGATCA	GAGAGCTCTGGTCAAGAGAT	GAAATGAGGTATGAGGAA		2758
RESULT 4						
AAF58655						
ID: AAF58655 standard; cDNA; 3420 BP.						

XX	AAF58655;
AC	
XX	27-APR-2001 (first entry)
DT	
XX	Human IB1 cDNA.
DE	
XX	Human; Repro-EN-1.0; IB1; cytostatic; breast cancer; uterine cancer;
KW	prostate cancer; epitore; s2
KW	

XX *Homo erectus* cf. *erectus*, ss.
OS *Homo sapiens*.

XX
PN W0200107616-A1.

XX
PD 01-FEB-2001.

XX
PF
10-MAR-2000; 2000WO-US06742.

PR 22-JUL-1999; 99US-0359084.

PR 30-JUL-1999; 99WO-US17284.
PR 23-NOV-1999; 99US-0447399.
...

XX
PA
...
(DIAG-) DIAGNOSTIC PROD CORP.

El Shami AS, Menon SN, French CK;

WPI; 2001-182795/18.

100

[illegible]

New autoantigens Repro-EN-1.0 and IB 1 polypeptides and polynucleotides are useful for diagnosing endometriosis or as a marker for pathologic conditions such as breast, uterine or prostate cancer

Claim 7; Page 28-34; 119pp; English.

The present sequence is given in a specification relating to recombinant polynucleotides comprising nucleotide sequences encoding a polypeptide epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a fully defined sequence of 860 and 937 amino acids. The epitope specifically binds to antibodies from subjects diagnosed with endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for diagnosing endometriosis by detecting antibodies in immunoassays, and are used as markers for pathologic conditions e.g. breast, uterine or prostate cancer. Methods for detecting conditions e.g. breast, uterine or polynucleotides or polypeptides are useful in the diagnosis of these cancers, monitoring their progress or treatment, and determining patient prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1 may be used as probes for detecting mRNA from cell types suspected to be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and IB 1 polypeptides and immunogenic fragments may be used as positive controls in diagnostic assays to detect antibodies that specifically bind the proteins from patient serum samples. The polypeptides are useful as immunogens for eliciting antibody production against epitopes of the protein, and as controls in diagnostic methods.

Sequence 3420 BP; 1036 A; 723 C; 778 G; 883 T; 0 other;

Query Match	90.4%	Score 2332.6	DB 22	Length 3420
Best Local Similarity	91.6%	Pred. No. 0		
Matches 2576	Conservative 0	Mismatches 4	Indels 231	Gaps 1
QY	1	ATGTCTCGGGTGGCTCTACCCACACTGTGTGGGACGTGAGAAAGTCCCTCGGG	60	
Db	176	ATGTCTCGGGTGGCTCTACCCACACTGTGTGGGACGTGAGAAAGTCCCTCGGG	235	
QY	61	CTGGAGGACCCGTCCTCCGGCTCGGAGTCGCTACCTGGGAAGACAGAAATTTATCCAAAGA	120	
Db	236	CTGGAGGACCCGTCCTCCGGCTCGGAGTCGCTACCTGGGAAGACAGAAATTTATCCAAAGA	295	
QY	121	TTAAACCTTGGAAGCAACCCCTTAATGTCATGATGGTTGTGTTTAATACAATCTCTTGAAT	180	
Db	296	TTAAACCTTGGAAGCAACCCCTTAATGTCATGATGGTTGTGTTTAATACAATCTCTTGAAT	355	
QY	181	GACACTGGAGAATATATTTATCTGGCTCAGATGACACCAATTTAGTAATAGTATCTCT	240	
Db	356	GACACTGGAGAATATATTTATCTGGCTCAGATGACACCAATTTAGTAATAGTATCTCT	415	
QY	241	TACAGCAGAAAGTTTGGACAAATTCGTTACGGCACCGAGCAACATATTTAGTGCA	300	
Db	416	TACAGCAGAAAGTTTGGACAAATTCGTTACGGCACCGAGCAACATATTTAGTGCA	475	
QY	301	AGGTTCTTACCTTGTACAAATGATAACAGATGTATCTGCTCTGGAGATGGAGTAATA	360	
Db	476	AGGTTCTTACCTTGTACAAATGATAACAGATGTATCTGCTCTGGAGATGGAGTAATA	535	
QY	361	TTTATACCAACGTTGACGAAAGATGAGAAACCAACAGCAATATTCAGTGTCAAT	420	
Db	536	TTTATACCAACGTTGACGAAAGATGAGAAACCAACAGCAATATTCAGTGTCAAT	595	
QY	421	TATGGAACCTACTTATGAGATATTAGCTGTACCAATGACCCCTTACACHTTCTCTCTGT	480	
Db	596	TATGGAACCTACTTATGAGATATTAGCTGTACCAATGACCCCTTACACHTTCTCTCTGT	655	
QY	481	GGTGAAGATGGAACTGTATAGTGGTTTATACACCGCATCAAACTPAGTCACAAAAGAA	540	
Db	656	GGTGAAGATGGAACTGTATAGTGGTTTATACACCGCATCAAACTPAGTCACAAAAGAA	715	
QY	541	GATTTGAAAGATGATATTTTAACTGTGCAGCTGCTGCCAGCTCTGTGCTATTGTC	600	
Db	716	GATTTGAAAGATGATATTTTAACTGTGCAGCTGCTGCCAGCTCTGTGCTATTGTC	775	
QY	601	CCACCAATACCATATATACCTTCCTGTTGGTTGTTCTGACAGCTCAGTACGAATATATGAT	660	

RESULT 5
 AAF26955
 ID AAF26955 standard; cDNA; 2448 BP.
 XX AC AAF26955;
 XX AC AAF26955;
 DT 10-APR-2001 (first entry)
 XX
 DE Human cancer associated antigen precursor HOM-TES-88/94/95 cDNA SEQ ID:8.
 XX Human; cancer associated antigen precursor; cancer associated antigen;
 KW seminoma; HLA; human leukocyte antigen; cytostatic; gene therapy;
 KW vaccine; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200100874-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-US17207.
 XX
 PR 30-JUN-1999; 99US-0346498.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Sahin U, Tureci O, Pfreundschuh M;
 XX
 DR WPI; 2001-112465/12.
 XX
 PT Diagnosing a disorder characterized by expression of a human cancer
 PT associated antigen precursor, comprises detecting interaction of an
 PT agent with a nucleic acid molecule encoding the antigen precursor
 PS
 XX Claim 48; Page 94; 126pp; English.
 XX
 CC The present invention describes a method for diagnosing a disorder
 CC characterised by expression of a human cancer associated antigen (CAA)
 CC precursor (I) coded by a NA Group 1 nucleic acid molecule (N1)
 CC comprising contacting the biological sample with an agent (A) that
 CC specifically binds to N1, (I) or its fragment, complexed with an human
 CC leukocyte antigen (HLA) molecule and determining the interaction between
 CC the agent and N1 or (I). (I) has cytostatic activity and can be used in
 CC gene therapy and vaccine production. The method can be used for treating
 CC a subject with a condition characterised by expression of (I) in cells
 CC of a subject. The present sequence represents a human cancer associated
 CC antigen precursor cDNA sequence which is used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 2448 BP; 778 A; 513 C; 551 G; 606 T; 0 other;

Query Match
 Best Local Similarity 82.9%; Score 2140; DB 22; Length 2448;
 Matches 2303; Conservative 0; Mismatches 45; Indels 66; Gaps 7;

QY 228 AATTAGTAACTCTTACAGCAGAAAGGTTTGACAAACAATTCGTCAGGSCACCGACCAA 287
 DB 16 AGTAAGGAGAGGTTTCTCATCAGAGTTTGACACAAATTCGTCAGGSCACCGACCAA 75
 QY 288 CATATTTAGTCAAAAGTTCTTACCTTTGACAAATGATAAACAAGATTGTATCTGCTGG 347
 DB 76 CATATTTAGTCAAAAGTTCTTACCTTTGACAAATGATAAACAAGATTGTATCTGCTGG 135
 QY 348 AGATGAGTAATATTTATACCAAGTTGAGCAAGATGCAGAAACCAACAGCAATGCCA 407
 DB 136 AGATGAGTAATATTTATACCAAGTTGAGCAAGATGCAGAAACCAACAGCAATGCCA 195
 QY 408 ATTACGTGTGATTGAACTACTTATGAGATTATGACTGTACCCATGACCCCTTACAC 467
 DB 196 ATTACGTGTGATTGAACTACTTATGAGATTATGACTGTACCCATGACCCCTTACAC 255

QY 468 TTTTCTCTCTGTGGTGAAGA-TGGAACCTGTAGGTGGTTTGATACACGCATCAAAACTA 526
 DB 256 TTTTCTCTCTGTGGTGAAGAGTGAACCTGTAGGTGGTTTGATACACGCATCAAAACTA 315
 QY 527 GCTGCACAAAAGAAGATTCTAAGATGATATTTAATTAAGTCTGACGTCGACGCT 586
 DB 316 GCTGCACAAAAGAAGATTCTAAGATGATATTTAATTAAGTCTGACGTCGACGCT 375
 QY 587 CTGTTT-GCTATTTGCCACCAATACCATATTA-CCTTCTGTGTGGTTCTTCTGACAGCT 643
 DB 376 CTGTTTGTCTATTTGCCACCAATACCATATTAACCTTCTGTGTGGTTCTTCTGACAGCT 435
 QY 644 CAGTACGATATATGATCGGCGAATGCTGGGCACAGAGCTACAGGGAATATTCAGGTC 703
 DB 436 CAGTACGATATATGATCGGCGAATGCTGGGCACAGAGCTACAGGGAATATTCAGGTC 495
 QY 704 GAGGACTACTGGAATGTTGCCCTTTTATTCCTTCCATCTTATATATAGTCTGCA 763
 DB 496 GAGGACTACTGGAATGTTG-CCTTTTATTCCTTCCATCTTATATATAGTCTGCA 554
 QY 764 GAGTACATCTCTGTGTACAGTGAAGTGTCAAGAGATTCCTGTAGTTACTCTTCAG 823
 DB 555 GAGTACATCTCTGTGTACAGTGAAGTGTCAAGAGATTCCTGTAGTTACTCTTCAG 614
 QY 824 ATTACATATATCTTTTTCACCCGAAAGATGATACAGCAGAGAACTTAAACTCCTTCG 883
 DB 615 ATTACATATATCTTTTTCACCCGAAAGATGATACAGCAGAGAACTTAAACTCCTTCG 674
 QY 884 CGAAGAGAGAGAGAGAGTGGACAAACACAGTTAAAGCTTTGAGACTTCGTGGTG 943
 DB 675 CGAAGAGAGAGAGAGAGTGGACAAACACAGTTAAAGCTTTGAGACTTCGTGGTG 734
 QY 944 ATTGCTCAGATCTGACCCAGAGAGAGCGGAGAGTGAACGAGAGAGAGAGAGAGC 1003
 DB 735 ATTGCTCAGATCTGACCCAGAGAGAGCGGAGAGTGAACGAGAGAGAGAGAGAGC 794
 QY 1004 AGATCCCAATGTCTCATTTGATGACAGAGATGCTGATATGTTTCAAGATGTTTGAAG 1063
 DB 795 AGATCCCAATGTCTCATTTGATGACAGAGATGCTGATATGTTTCAAGATGTTTGAAG 854
 QY 1064 AAGCAAGTGAAGTGGACAAAGCAATAGAGACGAGGAAGATCTCGACCCAGAGAGTGGAA 1123
 DB 855 AAGCAAGTGAAGTGGACAAAGCAATAGAGACGAGGAAGATCTCGACCCAGAGAGTGGAA 914
 QY 1124 CAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCATCAAGTCCCTGATTTGAAAGTGA 1183
 DB 915 CAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCATCAAGTCCCTGATTTGAAAGTGA 974
 QY 1184 GTGAACTGCAATGGAAGTAGATCTCCAGCTGAACAATTTCTCAGCCTTCTACATCCT 1243
 DB 975 GTGAACTGCAATGGAAGTAGATCTCCAGCTGAACAATTTCTCAGCCTTCTACATCCT 1034
 QY 1244 CTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACTC 1303
 DB 1035 CTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACTC 1094
 QY 1304 CTTTGCTATCTTCTCCAGACAGTGAACAAAGCAGTCTGTTGAGGATCTGGACACACA 1363
 DB 1095 CTTTGCTATCTTCTCCAGACAGTGAACAAAGCAGTCTGTTGAGGATCTGGACACACA 1154
 QY 1364 CACATCATCAGTCTGATAACAATAATGAAGCTGAGC----- 1401
 DB 1155 CACATCATCAGTCTGATAATCTCTTCTTCTGTTTAAACAACAGCTCGGATCCATGTCAC 1214
 QY 1402 -----CCGAACCCAGGACAGGTGAACCCAG 1426
 DB 1215 TTGACGACCAACAGGATACCATTAATGAAGCTGAGCCCAACCCAGGACAGGTGAACCCAG 1274
 QY 1427 TTTTAAAGTTTGACCTACAGCAGAGAGGAACAACTACAGCAACAATAAAGCTGAACCTTTA 1486
 DB 1275 TTTTAAAGTTTGACCTACAGCAGAGAGGAACAACTACAGCAACAATAAAGCTGAACCTTTA 1334
 QY 1487 CAGATGAATGGAGCAGTATACCATCAAGTTCTAGAGGAATGGGAGCCATTTGCAAACTCTG 1546

Db	1335	CAGATGAATGGACGAGTATAGCATCAAGTTC	AGAGAAATGGAGCCATTGCAAACTCTG	1394
Qy	1547	AGGTCAGGAGGAATCTTTCCGTCCACAGAGCTCAGTGC	CAACACACAGAAAGGACAGTGC	1606
Db	1395	AGGTCAGGAGGAATCTTTCCGTCCACAGAGCTCAGTGC	CAACACACAGAAAGGAGACAGTGC	1454
Qy	1607	AACAAAAGCTCCTGGAAGATCATCAGAGAGTGTGAC	CAAAATATCAGGAAGAGTATCTG	1666
Db	1455	AACAAAAGCTCCTGGAAGATCATCAGAGAGTGTGAC	CAAAATATCAGGAAGAGTATCTG	1514
Qy	1667	CAGAAAACCCAGTTTGAGAACCATATCAATATACAC	AAATCAGATPAAGTTCACAGCCAAAGC	1726
Db	1515	CAGAAAACCCAGTTTGAGAACCATATCAATATACAC	AAATCAGATPAAGTTCACAGCCAAAGC	1574
Qy	1727	CATTGGATTCAACTCAGGAGAAAGAAATGACCTCA	ATCTTGATCGCTCTGTGGGGTTC	1786
Db	1575	CATTGGATTCAACTCAGGAGAAAGAAATGACCTCA	ATCTTGATCGCTCTGTGGGGTTC	1634
Qy	1787	CAGAGAATCTGCTTCACTCTGAAAAGCCAAAGSAA	CCAGAAACTTCAGATCAGACTAGCA	1846
Db	1635	CAGAGAATCTGCTTCACTCTGAAAAGCCAAAGSAA	CCAGAAACTTCAGATCAGACTAGCA	1694
Qy	1847	CTGAGAGTCTACCAATGAAATACACCAATCCTGAG	CCCTCAGTTCACAAACAGAGCCA	1906
Db	1695	CTGAGAGTCTACCAATGAAATACACCAATCCTGAG	CCCTCAGTTCACAAACAGAGCCA	1754
Qy	1907	CTGGGCCCTTCAGTCTCATGAAGAAACATCCAC	CCAGGACTCTGCTTCTCAGGACACAGATG	1966
Db	1755	CTGGGCCCTTCAGTCTCATGAAGAAACATCCAC	CCAGGACTCTGCTTCTCAGGACAGATG	1813、
Qy	1967	ACAGTGAATGATGACCCGCTCTGATCCAGGTGCA	AGGTATCGAGCAGGACTCTGCTGATA	2036
Db	1814	ACAGTGAATGATGACCCGCTCTGATCCAGGTGCA	AGGTATCGAGCAGGACTCTGCTGATA	1870
Qy	2027	GACGCTCTGCTGTGCGCGTATTACGAGGTTC	TTCAGACGGAGAAAAGAAAGAAAGAAA	2086
Db	1871	GACGCTCTGCTGTGCGCGTATTACGAGGTTC	TTCAGACGGAGAAAAGAAAGAAAGAAA	1930
Qy	2087	TGGAAGAATTTGATACTTTGAACATTAAGACG	CGCTAGTATAAAGGCC	2146
Db	1931	TGGAAGAATTTGATACTTTGAACATTAAGACG	CGCTAGTATAAAGGCC	1990
Qy	2147	ATCGCAACTCCAGGCAATATGATAAAGAAC	CCAAATTCCTGGGGTGCCTAACCTTGTAAATGA	2206
Db	1991	ATCGCAACTCCAGGCAATATGATAAAGAAC	CCAAATTCCTGGGGTGCCTAACCTTGTAAATGA	2050
Qy	2207	GTGGTCTGACTGTGGCCACATTTTCATCTGGAT	CGGCACACTGCTGAGCATTGTATGTC	2266
Db	2051	GTGGTCTGACTGTGGCCACATTTTCATCTGGAT	CGGCACACTGCTGAGCATTGTATGTC	2110
Qy	2267	TTCTGGAAGCTGATAATCATGTGGTAAACTG	CGCTGCAGCCACATCCGTTTGACCCAAATTT	2326
Db	2111	TTCTGGAAGCTGATAATCATGTGGTAAACTG	CGCTGCAGCCACATCCGTTTGACCCAAATTT	2170
Qy	2327	TAGCCTCATCTGGCATAGATTATGACATAA	AGATCTGTCACCATTAGAAGAGTCAAGGA	2386
Db	2171	TAGCCTCATCTGGCATAGATTATGACATAA	AGATCTGTCACCATTAGAAGAGTCAAGGA	2230
Qy	2387	TTTTTTAACCGAAAACCTTGCTGATGAAGT	TATAACTCGAAAACGAACTCATGCTGGAAGAAA	2446
Db	2231	TTTTTTAACCGAAAACCTTGCTGATGAAGT	TATAACTCGAAAACGAACTCATGCTGGAAGAAA	2290
Qy	2447	CTAGAAAACACATTTACAGTTTCCAGGCTCT	TTTCATGTTGAGGATGTTGGCTTCACTTAATC	2506
Db	2291	CTAGAAAACACATTTACAGTTTCCAGGCTCT	TTTCATGTTGAGGATGTTGGCTTCACTTAATC	2350
Qy	2507	ATATCCGAGCTGACCGGTTGGAGGTCACAGAT	CAGAAAGGCTCTGCTCAAGAAATGAAA	2566
Db	2351	ATATCCGAGCTGACCGGTTGGAGGTCACAGAT	CAGAAAGGCTCTGCTCAAGAAATGAAA	2410
Qy	2567	ATGAGGATGAGGAA	2580	

Db	2411	ATGAGGATGAGGAA	2424
RESULT 6			
ABK35809			
ID	ABK35809	standard; cDNA: 1942 BP.	
XX	AC	ABK35809;	
XX	XX		
XX	DT	08-MAY-2002 (first entry)	
XX	XX		
XX	DE	cDNA sequence #200 encoding novel human secreted protein.	
XX	XX		
XX	KW	Human secreted protein; hyperproliferative disorder; autoimmune disorder;	
KW	KW	immune deficiency disorder; blood disorder; inflammatory disorder;	
KW	KW	infectious disorder; allergic condition; neurodegenerative disorder;	
KW	KW	liver fibrosis; coagulation disorder; gene therapy; antimicrobial;	
KW	KW	tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; sa;	
XX	OS	Homo sapiens.	
OS	XX		
PN	PN	WO200117289-A2.	
XX	XX	18-OCT-2001.	
XX	XX		
XX	PF	29-MAR-2001; 2001WO-US10232.	
XX	XX		
XX	PR	06-APR-2000; 2000US-195605P.	
XX	XX	(GEMY) GENETICS INST INC.	
XX	XX	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;	
PI	PI	Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;	
PI	PI	Clark HF, Fachtel K, Howes SH, Resnick RJ, Galukota K, Graham JR;	
XX	XX		
DR	XX	WPI; 2002-179322/23.	
XX	XX		
XX	PT	Six hundred and twenty three polynucleotides derived from a variety of	
PT	PT	human tissue sources which encode secreted proteins, useful for	
PT	PT	treating immune deficiencies and disorders such as autoimmune disorders	
PT	PT		
XX	XX		
PS	XX	Claim 1; Page 194-195; 393pp; English.	
XX	XX		
CC	CC	The present invention relates to the isolation of novel cDNA sequences	
CC	CC	which encode human secreted proteins. The cDNA sequences have been	
CC	CC	derived from a variety of human tissues. The invention also provides	
CC	CC	a method for producing proteins from these polynucleotide sequences.	
CC	CC	The proteins are useful for identifying compounds that modulate their	
CC	CC	activity and production. The sequences of the invention are	
CC	CC	useful for treating diseases such as hyperproliferative disorders	
CC	CC	(e.g. cancer), immune deficiency disorders (e.g. severe combined	
CC	CC	immunodeficiency (SCID)), autoimmune disorders (e.g. multiple	
CC	CC	sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory	
CC	CC	disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),	
CC	CC	allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.	
CC	CC	Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.	
CC	CC	haemophilia), and tumours. The polynucleotide sequences of the	
CC	CC	invention are also useful in gene therapy. ABK35610-ABK3232 represent	
CC	CC	the cDNA sequences of the invention that encode for novel human	
CC	CC	secreted proteins.	
XX	XX		
SQ	XX	Sequence 1942 BP; 626 A; 383 C; 420 G; 513 T; 0 other;	
		Query Match	59.8%; Score 1542.4; DB 24; Length 1942;
		Best Local Similarity	99.9%; Pred. No. 0;
		Matches 1543; Conservative	0; Mismatches 1; Indels 0; Gaps

QY	1037	CTGATATGTTTATCAAGATGCTTTTGAAAGCAAGTAGTGAGGTTGCACAAAGCAATAGAGGAC	1096
DB	1	CTGATATGTTTATCAAGATGCTTTTGAAAGCAAGTAGTGAGGTTGCACAAAGCAATAGAGGAC	60
QY	1097	GAGAAGATCTGACCCAGGATGGGACAGTCAATCAGATATTTCACTCTTCCTACGG	1156

Db	1141	COAATTTCTGGGGTGTAACATTTGTAATGAGTGGTTCTGACTGTGCCACATTTTCATCT	1200
Qy	2237	GGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCATGTGGTAAACT	2296
Db	1201	GGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCATGTGGTAACT	1260
Qy	2297	GCCTGACGCCACATCCGTTTGACCCAAATTTAGCCCTCATCTGGCATAGATATGACATAA	2356
Db	1261	GCCTGACGCCACATCCGTTTGACCCAAATTTAGCCCTCATCTGGCATAGATATGACATAA	1320
Qy	2357	AGATCTGCTGACCATTAGAAGAGTCAAGGATTTTAAACGAAAACCTTGCTGATGAAGTTA	2416
Db	1321	AGATCTGCTGACCATTAGAAGAGTCAAGGATTTTAAACGAAAACCTTGCTGATGAAGTTA	1380
Qy	2417	TAACTCGAAACGAACTCATGCTGGGAAGAACTAGAAACACCATTCAGTTCAGGCTCTT	2476
Db	1381	TAACTCGAAACGAACTCATGCTGGGAAGAACTAGAAACACCATTCAGTTCAGGCTCTT	1440
Qy	2477	TCATGTTGAGGATGTTGGCTTTCACTTAATCATATCCGAGCTGACCGGTTGAGGGGTGACA	2536
Db	1441	TCATGTTGAGGATGTTGGCTTTCACTTAATCATATCCGAGCTGACCGGTTGAGGGGTGACA	1500
Qy	2537	GATCAGAAGGCTCTGGTCAAGAGAAATGAAATGAGGATGAGGAA	2580
Db	1501	GATCAGAAGGCTCTGGTCAAGAGAAATGAAATGAGGATGAGGAA	1544

RESULT 7
ABA93744
ID ABA93744 standard; cDNA; 1952 BP.
XX
AC ABA93744;
XX
DT 30-APR-2002 (first entry)
XX
XX Human signal transduction cDNA clone tes3_11c22.
DE
DE Human; foetal brain; foetal kidney; melanoma; testis; amygdala;
KW gene therapy; chromosome 1q23.2-24.3; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200198454-A2.
PN
XX
XX 27-DEC-2001.
PD
XX
XX 25-APR-2001; 2001WO-IB02050.
PF
XX
XX 25-APR-2000; 2000US-199380P.
PPR
XX
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
FPA
XX
XX Wiemann S;
PI
XX
XX WPI; 2002-055860/07.
DR
XX
XX P-PSDB; ABB05707.
TT
XX
XX Human cDNA sequences and clones derived from human fetal brain, fetal
XX kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
XX screening and therapy -
XX
XX Claim 1; Page 285-286; 61pp; English.
XX
XX The present invention describes assemblages and computer readable media
XX comprising novel human cDNA sequences and clones derived from human
XX foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
XX libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
XX present invention which encode the proteins given in ABB05662 to
XX ABB05729. The human cDNA sequences and clones can be used in gene
XX therapy. The clones may be used in a variety of applications, for
XX example they may be used in profiling assays, for providing large arrays
XX of human genetic material for implementing large-scale screening
XX strategies and for treating diseases via gene therapy procedures.

XX	SQ	Sequence	1952 BP	642 A	392 C	418 G	500 T	0 other	Query Match	56.1%	Score	1446.4	DB	24	Length	1952
									Best Local Similarity	96.1%	Pred. No.	0				
									Mismatches	0	Mismatches	1			Indels	60
									Matches	1517	Conservative					Gaps
QY	1063	GAACCAAGTGGGTTGCACAAAGCAATATAGAGGACGAGGAGAAGATCTCGACCCAGAGGTGGA	1122													
Db	1	GAACCAAGTGGGTTGCACAAAGCAATATAGAGGACGAGGAGAAGATCTCGACCCAGAGGTGGA	60													
QY	1123	ACAAAGTCAATCAGATATTTCAACTTTCTTACGGTCCCATCAAGTCTCTGATTTGGAGTG	1182													
Db	61	ACAAAGTCAATCAGATATTTCAACTTTCTTACGGTCCCATCAAGTCTCTGATTTGGAGTG	120													
QY	1183	AGTGAAGTCAATGGAAGTATACCTCCAGCTGAACAAATTTCTTACGCTCTCTACATCC	1242													
Db	121	AGTGAAGTCAATGGAAGTATACCTCCAGCTGAACAAATTTCTTACGCTCTCTACATCC	180													
QY	1243	TCTACAATGTTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTCACT	1302													
Db	181	TCTACAATGTTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTCACT	240													
QY	1303	CGTTTGTCTATCTTCTCCACAGAGTGAACAAAGGACGCTGTGTGAGGCATCTGGACACAC	1362													
Db	241	CGTTTGTCTATCTTCTCCACAGAGTGAACAAAGGACGCTGTGTGAGGCATCTGGACACAC	300													
QY	1363	ACACATCATCTAGTCT	1377													
Db	301	ACACATCATCTAGTCTGATCTCTCTCTGTGGTTAAACAAACAGCTCGGATCCATGTCA	360													
QY	1378	-----GATAACAAATAATGAAAGCTGAGCCCCAAACACGAGGACAGGTGAA	1422													
Db	361	CTTGACGACCAACAGGATTAACAATAATGAAAGCTGAGCCCCAAACACGAGGACAGGTGAA	420													
QY	1423	CCAGTTTTAAGTTTGCATCTACAGCACAGNAGAAACAACTTACAAGCACAAATAAAACTGAAC	1482													
Db	421	CCAGTTTTAAGTTTGCATCTACAGCACAGNAGAAACAACTTACAAGCACAAATAAAACTGAAC	480													
QY	1483	TTTACAGATGMAATGGACGAGTATAGCATCAAGTTCTTAGAGGAATTTGGAGCCATTTGCAAA	1542													
Db	481	TTTACAGATGMAATGGAGCAGTATAGCATCAAGTTCTTAGAGGAATTTGGAGCCATTTGCAAA	540													
QY	1543	CTGAGGGTTCAGAGGAATCTTTTCGTCCACAGAGCTCAGTCCCAACACCAAGAGGAC	1602													
Db	541	CTGAGGGTTCAGAGGAATCTTTTCGTCCACAGAGCTCAGTCCCAACACCAAGAGGAC	600													
QY	1603	AGTGAACAAAGGCTCCTGAAGAAATCATCAGAGGATGTGACAAAAATATCAGGAAGGAT	1662													
Db	601	AGTGAACAAAGGCTCCTGAAGAAATCATCAGAGGATGTGACAAAAATATCAGGAAGGAT	660													
QY	1663	CTGCGAGAAACCCAGTTTGAGAACCATATCAATATAACACAACTCAGATAGTTTCACGCC	1722													
Db	661	CTGCGAGAAACCCAGTTTGAGAACCATATCAATATAACACAACTCAGATAGTTTCACGCC	720													
QY	1723	AAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTTGTGGG	1782													
Db	721	AAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTTGTGGG	780													
QY	1783	GTTCGAGAGAAATCTGCTTCAATGTAAAAAGCCAGGAACCAAGAACTTCAGATCAGACT	1842													
Db	781	GTTCGAGAGAAATCTGCTTCAATGTAAAAAGCCAGGAACCAAGAACTTCAGATCAGACT	840													
QY	1843	AGCACTGAGAGTGTACCAATGAAATTAACCAAACTCCTGAGCCTCAGTTTCCAAACAGAA	1902													
Db	841	AGCACTGAGAGTGTACCAATGAAATTAACCAAACTCCTGAGCCTCAGTTTCCAAACAGAA	900													
QY	1903	GCCACTGGGCCCTTCAGCTCATGAGAAACATCCACAGGACCTCTGCTCTTTCAGGACACA	1962													
Db	901	GCCACTGGGCCCTTCAGCTCATGAGAAACATCCACAGGACCTCTGCTCTTTCAGGACACA														

Db	961	GATCACAGTGATGATGACCCAGTCTCTGATCCAGTGCAAGGTATCGACGACGACCTGGT	1022
Qy	2023	GATGACGCTCTGCTGTTGGCCGATTCAGGAGTTCTTCAGACGGAGAGAAAGAAAGAAA	2082
Db	1021	GATGACGCTCTGCTGTTGGCCGATTCAGGAGTTCTTCAGACGGAGAGAAAGAAAGAAA	1080
Qy	2083	GAATGGAAGAAATTCGATCTACTTTGAACATTTAGAAGCCCTAGTAAAAATGTTTATAAA	2142
Db	1081	GAATGGAAGAAATTCGATCTACTTTGAACATTTAGAAGCCCTAGTAAAAATGTTTATAAA	1140
Qy	2143	GGCCATCGCACTCCAGGACCAATGATAAAGAGCCAAATTTCTGGGGTCTAACTTTTGA	2202
Db	1141	GGCCATCGCACTCCAGGACCAATGATAAAGAGCCAAATTTCTGGGGTCTAACTTTTGA	1200
Qy	2203	ATGAGTGGTTCCTGACTGTGGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTG	2262
Db	1201	ATGAGTGGTTCCTGACTGTGGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTG	1260
Qy	2263	ATGCTTCTGGAAGCTGATTAATCATGTGTGTAATACTGCTGCAGCCACATCGTTTGACCCA	2322
Db	1261	ATGCTTCTGGAAGCTGATTAATCATGTGTGTAATACTGCTGCAGCCACATCGTTTGACCCA	1320
Qy	2323	ATTTTACGCTCATCTCGGCATAGATTATGACATAAAGATCTGTCACCATTTAGAAGAGTCA	2382
Db	1321	ATTTTACGCTCATCTCGGCATAGATTATGACATAAAGATCTGTCACCATTTAGAAGAGTCA	1380
Qy	2383	AGGATTTTTAACCGAAACTTCTGATGAAGTTTAACTCGAAACGAACTCATGCTGGAA	2442
Db	1381	AGGATTTTTAACCGAAACTTCTGATGAAGTTTAACTCGAAACGAACTCATGCTGGAA	1440
Qy	2443	GAACATAGAAACACCACTTACAGTTCACAGCCTCTTTCATGTTGAGGATGTTGGCTTCACCT	2502
Db	1441	GAACATAGAAACACCACTTACAGTTCACAGCCTCTTTCATGTTGAGGATGTTGGCTTCACCT	1500
Qy	2503	AATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGGCTCTGGTCAAGAGAAT	2562
Db	1501	AATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGGCTCTGGTCAAGAGAAT	1560
Qy	2563	GAATAAGGAGATGAGGAA 2580	
Db	1561	GAATAAGGAGATGAGGAA 1578	
RESULT 8			
AAH13776			
ID	AAH13776 standard; cdna; 1726 BP.		
XX	AAH13776;		
XX	26-JUN-2001 (first entry)		
XX	Human cdna sequence SEQ ID NO:10702.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy;		
XX	Homo sapiens.		
XX	EP1074617-A2.		
XX	07-FEB-2001.		
XX	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
XX	27-AUG-1999; 99JP-0300253.		
XX	11-JAN-2000; 2000JP-0118776.		
XX	02-MAY-2000; 2000JP-0183767.		
XX	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T		

Best Local Similarity 95.1%; Pred. No. 3.6e-218;
Matches 851; Conservative 0; Mismatches 2; Indels 42; Gaps 1

24-FEB-2000; 2000US-0184772.
24-FEB-2000; 2000US-0184773.
24-FEB-2000; 2000US-0184774.
24-FEB-2000; 2000US-0184776.
24-FEB-2000; 2000US-0184777.
24-FEB-2000; 2000US-0184797.
24-FEB-2000; 2000US-0184813.
24-FEB-2000; 2000US-0184837.
24-FEB-2000; 2000US-0184841.
24-FEB-2000; 2000US-0185213.
24-FEB-2000; 2000US-0185216.
12-MAY-2000; 2000US-0203785.
15-MAY-2000; 2000US-0204226.
16-MAY-2000; 2000US-0204525.
16-MAY-2000; 2000US-0204821.
16-MAY-2000; 2000US-0204908.
16-MAY-2000; 2000US-0205232.
17-MAY-2000; 2000US-0204815.
17-MAY-2000; 2000US-0204863.
17-MAY-2000; 2000US-0205221.
17-MAY-2000; 2000US-0205285.
17-MAY-2000; 2000US-0205286.
17-MAY-2000; 2000US-0205287.
17-MAY-2000; 2000US-0205323.
17-MAY-2000; 2000US-0205324.
(INCY-) INCYTE GENOMICS INC.

Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
Chen A, D'Sa SA, Amshery S, Dahl CR, Dam TC, Daniels SE;
Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
WPI: 2001-502867/55.
P-PSDB: AAU19599.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
enzymes, hormones and receptors, useful in diagnostics and therapeutics

Claim 1; Page 384; 522pp; English.

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DTHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DTHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DTHPs, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DTHPs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies against DTHPs and in assays to identify modulators of DTHP expression and activity. The anti-DTHP antibodies and antagonists may also be used to down regulate expression and activity. The anti-DTHP antibodies may also be used as diagnostic agents for detecting the presence of DTHPs in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic (DTHP) polynucleotides of the invention.

Sequence 1012 BP; 366 A; 198 C; 221 G; 226 T; 1 other;

24-FEB-2000; 2000US-0184772.
24-FEB-2000; 2000US-0184773.
24-FEB-2000; 2000US-0184774.
24-FEB-2000; 2000US-0184776.
24-FEB-2000; 2000US-0184777.
24-FEB-2000; 2000US-0184797.
24-FEB-2000; 2000US-0184813.
24-FEB-2000; 2000US-0184837.
24-FEB-2000; 2000US-0184841.
24-FEB-2000; 2000US-0185213.
24-FEB-2000; 2000US-0185216.
12-MAY-2000; 2000US-0203785.
15-MAY-2000; 2000US-0204226.
16-MAY-2000; 2000US-0204525.
16-MAY-2000; 2000US-0204821.
16-MAY-2000; 2000US-0204908.
16-MAY-2000; 2000US-0205232.
17-MAY-2000; 2000US-0204815.
17-MAY-2000; 2000US-0204863.
17-MAY-2000; 2000US-0205221.
17-MAY-2000; 2000US-0205285.
17-MAY-2000; 2000US-0205286.
17-MAY-2000; 2000US-0205287.
17-MAY-2000; 2000US-0205323.
17-MAY-2000; 2000US-0205324.
(INCY-) INCYTE GENOMICS INC.

Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
Chen A, D'Sa SA, Amshery S, Dahl CR, Dam TC, Daniels SE;
Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
WPI: 2001-502867/55.
P-PSDB: AAU19599.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
enzymes, hormones and receptors, useful in diagnostics and therapeutics

Claim 1; Page 384; 522pp; English.

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DTHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DTHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DTHPs, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DTHPs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies against DTHPs and in assays to identify modulators of DTHP expression and activity. The anti-DTHP antibodies and antagonists may also be used to down regulate expression and activity. The anti-DTHP antibodies may also be used as diagnostic agents for detecting the presence of DTHPs in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic (DTHP) polynucleotides of the invention.

Sequence 1012 BP; 366 A; 198 C; 221 G; 226 T; 1 other;

24-FEB-2000; 2000US-0184772.
24-FEB-2000; 2000US-0184773.
24-FEB-2000; 2000US-0184774.
24-FEB-2000; 2000US-0184776.
24-FEB-2000; 2000US-0184777.
24-FEB-2000; 2000US-0184797.
24-FEB-2000; 2000US-0184813.
24-FEB-2000; 2000US-0184837.
24-FEB-2000; 2000US-0184841.
24-FEB-2000; 2000US-0185213.
24-FEB-2000; 2000US-0185216.
12-MAY-2000; 2000US-0203785.
15-MAY-2000; 2000US-0204226.
16-MAY-2000; 2000US-0204525.
16-MAY-2000; 2000US-0204821.
16-MAY-2000; 2000US-0204908.
16-MAY-2000; 2000US-0205232.
17-MAY-2000; 2000US-0204815.
17-MAY-2000; 2000US-0204863.
17-MAY-2000; 2000US-0205221.
17-MAY-2000; 2000US-0205285.
17-MAY-2000; 2000US-0205286.
17-MAY-2000; 2000US-0205287.
17-MAY-2000; 2000US-0205323.
17-MAY-2000; 2000US-0205324.
(INCY-) INCYTE GENOMICS INC.

Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
Chen A, D'Sa SA, Amshery S, Dahl CR, Dam TC, Daniels SE;
Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffo A;
Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
WPI: 2001-502867/55.
P-PSDB: AAU19599.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
enzymes, hormones and receptors, useful in diagnostics and therapeutics

Claim 1; Page 384; 522pp; English.

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DTHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DTHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DTHPs, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DTHPs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies against DTHPs and in assays to identify modulators of DTHP expression and activity. The anti-DTHP antibodies and antagonists may also be used to down regulate expression and activity. The anti-DTHP antibodies may also be used as diagnostic agents for detecting the presence of DTHPs in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic (DTHP) polynucleotides of the invention.

Sequence 1012 BP; 366

XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
 KW Homo sapiens.
 OS WO200166719-A1.
 PN XX
 PD 13-SEP-2001.
 XX
 PF 02-MAR-2001; 2001WO-JP01629.
 XX
 PR 07-MAR-2000; 2000JP-0159195.
 XX
 PA (CHIB-) CHIBA PREFECTURE.
 PA (HISM) HISAMITSU PHARM CO LTD.
 XX
 PI Nakagawara A;
 DR WPI; 2001-565584/63.
 XX
 XX Nucleic acids originating in gene expressed in human neuroblastoma,
 PT useful as probe or primer in diagnosing prognosis of human
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker
 PT for anti-cancer agents
 XX
 PS Claim 1; Page 2858; 2979pp; Japanese.
 XX
 CC The invention relates to novel genes (AAI93926-AAI97963) expressed in
 CC human neuroblastoma. The nucleic acids are applicable as a probe or
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
 CC susceptibility indicators or tumour markers for anti-cancer agents. The
 CC gene information for diagnosing prognosis is related to factors similar
 CC to that for N-myc and TrkA genes.
 XX
 SQ Sequence 728 BP; 231 A; 159 C; 169 G; 167 T; 2 other;
 Query Match 26.0%; Score 671.4; DB 22; Length 728;
 Best Local Similarity 99.6%; Pred. No. 7.9e-182;
 Matches 694; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 830 TATATCTTTTGGCCGAAAGATGATACGACGAGAGACTTAAACCTCTCTCGCGAAG 91
 DB 32 TATATCTTTTGGCCGAAAGATGATACGACGAGAGACTTAAACCTCTCTCGCGAAG 91
 QY 890 AG 949
 DB 92 AG 151
 QY 950 CAGATAGTGGAG 1009
 DB 152 CAGATAGTGGAG 211
 QY 1010 CCAATGTGTCTTGTATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1069
 DB 212 CCAATGTGTCTTGTATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
 QY 1070 GTGAGGTGTCACAAAGCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1129
 DB 272 GTGAGGTGTCACAAAGCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331
 QY 1130 AATCAGATATTTCAACTCTTCTACGCTCCCAATCAGTCTGATTTGAGAGAGAGAG 1189
 DB 332 AATCAGATATTTCAACTCTTCTACGCTCCCAATCAGTCTGATTTGAGAGAGAGAG 391
 QY 1190 CTGCAATGGAAGTAGATCTCCAGCTGACAAATTTCTTCAGCTCTTCTACATCTCTAC 1249
 DB 392 CTGCAATGGAAGTAGATCTCCAGCTGACAAATTTCTTCAGCTCTTCTACATCTCTAC 451
 QY 1250 TGTACGCTCAGGCTCATTCGACATCATCTCCACAGAGAGAGAGAGAGAGAGAGAG 1309
 DB 452 TGTACGCTCAGGCTCATTCGACATCATCTCCACAGAGAGAGAGAGAGAGAGAGAG 511
 QY 1310 TATCTTCTCCAG 1369

DB 512 TATCTTCTCAGACAGTGAACAAAGCAGTCTCTGTGAGCATCTGACACCCACACATC 571
 QY 1370 ATCAGTCTGATACAAATATGAAAGCTGAGCCCCAAACAGGAGAGAGTGAACAGTTT 1429
 DB 572 ATCAGTCTGATACAAATATGAAAGCTGAGCCCCAAACAGGAGAGTGAACAGTTT 631
 QY 1430 TAAGTTTGCACTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
 DB 632 TAAGTTTGCACTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691
 QY 1489 GATGAAT-GGAGCAGTATAGCATCAAGTTCTAGAGGA 1524
 DB 692 GATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGA 728
 RESULT 11
 AAA44677
 ID AAA44677 standard; cDNA; 635 BP.
 XX
 AC AAA44677;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:1252.
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
 KW antitumor; osteoprotective; neuroprotective; nontropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX
 PR 15-OCT-1998; 98US-0104436.
 XX
 XX (GEMY) GENETICS INST INC.
 PA
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 XX
 XX WPI; 2000-317938/27.
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (sESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 PS Claim 1; Page 527; 803pp; English.
 XX
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The sESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antidiabetic;
 CC cytostatic; antibacterial; antifungal; antitumor; antipsoriatic;
 CC antiasthmatic; vulnerary; antitumor; osteoprotective; neuroprotective;
 CC nontropic; antiparkinsonian; antipsoriatic; cerebroprotective;

Probe #12432 for gene expression analysis in human heart cell sample.

Human; gene expression; heart; microarray; vascular system; probe; cardiovascular disease; hypertension; cardiac arrhythmia; congenital heart disease; ss.

xx Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

XY AC

301 CACCTATCCACGACCTGGTGATG 328

XY AC

AAI21817	
ID	AAI21817 standard; DNA; 332 Bp.
XX	
AC	AAI21817;
XX	
DT	12-OCT-2001 (first entry)
XX	
DE	Probe #11750 for gene expression analysis in human cervical cell sample.
XX	
KW	Probe; human; microarray; gene expression; cervical epithelial cell;
KW	cervical cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157278-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00670.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
PY	

04 OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.
Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -
Claim 25; SEQ ID No 11750; 487pp; English.

The present invention relates to human single exon nucleic acid probes (SENP). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed specification, was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 332 BP; 107 A; 84 C; 74 G; 67 T; 0 other;

Query Match 12.7%; Score 328; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.5e-83;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

b
 181 CTGACCTCAGTTCACACAGAGCCACTGGGCTTCAGTCCATGAAGAAACATCCACCA 240
 1940 GGGACTCGCTCTCAGGACACAGATGACAGTGCATGATGACCCAGTCCTGATCCAGGTG 1999

[illegible]

RESULT 19	
ABSI15038	
ID	ABSI15038 standard; DNA; 332 BP.
XX	
XX	
AC	
ABSI15038;	
XX	
XX	
DT	19-AUG-2002 (first entry)
XX	
DE	Human genome-derived single exon probe ORF from lung SEQ ID NO 15029.
DE	
XX	
XX	Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW	chronic obstructive pulmonary disease; interstitial lung disease;
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW	tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW	pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW	primary ciliary dyskinesia; pulmonary hypertension;
KW	lung disease; frame: ORF.

Homo sapiens.
WO200186003-A2.
15-NOV-2001.
30-JAN-2001; 2001WO-US00665.
04-FEB-2000; 2000US-180312P.
26-MAY-2000; 2000US-207456P.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-234687P.
27-SEP-2000; 2000US-236359P.
04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
wpt. 2002-114183/15.

spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples - English.

Claim 4; SEQ ID NO 15029; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of

Accession	Sequence	Position
db		241
QY	GGGACTGCTCTCAGGACACAGATGACGTGATGATGACCCAGTCTTGATCCCGTG	300
		2000
	CAAGGTATCGAGCAGGACCTGGTGATG	2027
		338
		328

RESULT 18
AAI07503
ID AAI07503 standard; DNA; 332 BP.
XX
AC
AAI07503;
XX
DT
09-OCT-2001 (first entry)
XX
DE
Probe #7494 used to measure gene expression in human breast sample.
XX
DE
probe: human; breast disease; breast cancer; development disorder; ss;
XX
KW
eco;cancer;carcinoma tumour.

X	Homo sapiens.	
S		
X	WO200157270-A2.	
X		
N		
N		
X		
X		
D	09-AUG-2001.	
X		
X		
F	29-JAN-2001; 2001WO-US000661.	
F		
X		
X		
R	04-FEB-2000; 2000US-0180312.	
R	26-MAY-2000; 2000US-0207456.	
R	30-JUN-2000; 2000US-0608408.	
R	03-AUG-2000; 2000US-0632366.	
R	21-SEP-2000; 2000US-0234687.	
R	21-SEP-2000; 2000US-0236359.	
R	27-SEP-2000; 2000US-0236359.	
R	04-OCT-2000; 2000GB-0024263.	
XX		
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
XX		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
PI		
PI		
XX		
XX	WPI; 2001-476286/51.	
XX		
XX	Novel single exon nucleic acid probe used to measuring gene expression	
PT		
PT	in a human breast -	
XX		
XX		

Claim 25: SEQ ID NO 743, 752-777

xx The present invention relates to novel single exon nucleic acid probes.
xx The present sequence is one such probe. The probes are useful for
cc measuring human gene expression in a human breast sample, where the probe
cc hybridises at high stringency to a nucleic acid expressed in the human
cc breast. The probes are useful for predicting, diagnosing, grading,
cc staging, monitoring and prognosing diseases of the human breast,
cc particularly those diseases with polygenic aetiology. The diseases
cc include: breast cancer, disorders of development, inflammatory diseases
cc of the breast, fibrocystic changes, proliferative breast disease and
cc non-carcinoma tumours.
cc Note: The sequence data for this patent did not form part of the printed
cc specification, but was obtained in electronic format directly from WIPO
cc database sequences.

CC at ftp.wipo.int/pub/published_publications
XX Sequence 332 BP; 107 A; 84 C; 74 G; 67 T; 0 other;
SQ
Query Match 12.7%; Score 328; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.5e-83;
Gaps 0; Indels 0;

	Matches	328;	Conservative	0;	Mismatches	0;	Indels	0;	GC%
QY	1700	CACAAATCAGATTAAGTTCACAGCAAGCCATTGGATTCCAATTCAGGAGAAAGAATGACC	1759						
Db	1	CACAAATCAGATTAAGTTCACAGCAAGCCATTGGATTCCAATTCAGGAGAAAGAATGACC	60						
QY	1760	TCAATCTTGATCGCTCTTTGTGGGTTCCAGAGAATCTGCTTCATCTGAAAAGGCAAGG	1819						

CC probes: the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemolysis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagen syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIP0 at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 332 BP; 107 A; 84 C; 74 G; 67 T; 0 other;

Query Match	12.7%;	Score 328;	DB 24;	Length 332;
Best Local Similarity	100.0%;	Pred. No. 1.5e-83;		
Matches 328;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1700 CACATCAGATAGTTTCACAGCCAGCCATTGGATTCCCACTCAGGAGAAATGACC 1759
 DB 1 CACATCAGATAGTTTCACAGCCAGCCATTGGATTCCCACTCAGGAGAAATGACC 60
 QY 1760 TCAATCTGTGCTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAAGCCAAAG 1819
 DB 61 TCAATCTGTGCTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAAGCCAAAG 120
 QY 1820 AACAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACCAATC 1879
 DB 121 AACAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACCAATC 180
 QY 1880 CTGAGCCTCAGTCCCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAATCAACCA 1939
 DB 181 CTGAGCCTCAGTCCCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAATCAACCA 240
 QY 1940 GGGACTGTGCTCTTCAGGACACAGATGACAGTGTGATGATGATGATGATGATGATG 1999
 DB 241 GGGACTGTGCTCTTCAGGACACAGATGACAGTGTGATGATGATGATGATGATGATG 300
 QY 2000 CAAGGTATCCAGGACGACCTGGTGATG 2027
 DB 301 CAAGGTATCCAGGACGACCTGGTGATG 328

RESULT 20
 AA197812/c
 ID AA197812 standard; cDNA; 710 BP.
 XX AA197812;
 AC
 XX DT 13-NOV-2001 (first entry)

XX	Human neuroblastoma expressed polynucleotide SEQ ID NO 3887.
DE	Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX	Homo sapiens.
XX	WO200166719-A1.
XX	13-SEP-2001.
XX	02-MAR-2001; 2001WO-JP01629.
XX	07-MAR-2000; 2000JP-0159195.
XX	(CHIB-) CHIBA PREFECTURE.
XX	(HISM) HISAMITSU PHARM CO LTD.
XX	Nakagawara A;
XX	WPI; 2001-565584/63.
XX	Nucleic acids originating in gene expressed in human neuroblastoma,
XX	useful as probe or primer in diagnosing prognosis of human
XX	neuroblastoma, malignancy and susceptibility indicator or tumour marker
XX	for anti-cancer agents -
XX	Claim 1; Page 2859; 2979pp; Japanese.
XX	The invention relates to novel genes (AA193926-AA197963) expressed in
XX	human neuroblastoma. The nucleic acids are applicable as a probe or
XX	primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX	susceptibility indicators or tumour markers for anti-cancer agents. The
XX	gene information for diagnosing prognosis is related to factors similar
XX	to that for N-myc and TrkA genes.
XX	Sequence 710 BP; 218 A; 139 C; 109 G; 240 T; 4 other;

Query Match	11.2%;	Score 287.8;	DB 22;	Length 710;
Best Local Similarity	98.4%;	Pred. No. 7.8e-72;		
Matches 300;	Conservative 0;	Mismatches 4;	Indels 1;	Gaps 1;

QY 2277 TGATTAATCATGTTGGTAAACTGCCTGCAGCC-ACATCCGTTTGCACCAATTTTAGCCTCAT 2335
 DB 699 TAATTAATCATGTTGGTAAACTGCCTGCAGCCACACATCCGTTTGCACCAATTTTAGCCTCAT 640
 QY 2336 CTGGCATAGATTATGACATARAAGATCTGGTCACCATTAAGAGAGTCAAGGATTTTAAACC 2395
 DB 639 CTGGCATAGATTATGACATARAAGATCTGGTCACCATTAAGAGAGTCAAGGATTTTAAACC 580
 QY 2396 GAAAACCTTGCTGATGAAGTTATACTCGAAACGAACTCATGCTGGAAGAACTAGAAACA 2455
 DB 579 GAAAACCTTGCTGATGAAGTTATACTCGAAACGAACTCATGCTGGAAGAACTAGAAACA 520
 QY 2456 CCATTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAG 2515
 DB 519 CCATTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAG 460
 QY 2516 CTGACCGGTTGGAGGTTACAGATCAGAGGCTCTGCTGCAAGAGAAATGAAGATGAGGATG 2575
 DB 459 CTGACCGGTTGGAGGTTACAGATCAGAGGCTCTGCTGCAAGAGAAATGAAGATGAGGATG 400
 QY 2576 AGGAA 2580
 DB 399 AGGAA 395

RESULT 21
 ABA46979
 ID ABA46979 standard; DNA; 261 BP.
 XX
 AC ABA46979;
 XX

us-09-781-693a-1.rng

Wed Mar 26 09:38:49 2003

```

DB 181 CTACTCTTTGCTATCTTCCAGACAGTGACAAAGGAGCTGTGTGGCATCTGGAC 240
QY 1358 ACCACACACATCATCATCTG 1378
DB 241 ACCACACACATCATCATCTG 261

RESULT 22
ID ABA64860 standard; DNA; 261 BP.
XX AC ABA64860;
XX AC
XX 01-FEB-2002 (first entry)
XX 01-FEB-2002 (first entry)
XX Human foetal liver single exon nucleic acid probe #13165.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
XX WO200157277-A2.
XX PD
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 4; SEQ ID NO 13165; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;
XX Query Match 10.1%; Score 261; DB 22; Length 261;
XX Best Local Similarity 100.0%; Pred. No. 2.2e-64;
XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTGG 1177
DB 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTGG 60
QY 1178 AAGTGAGTGAACACTGCAATGGAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237
DB 61 AAGTGAGTGAACACTGCAATGGAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120
QY 1238 CATCTCTTCAATGTGCTCAGCTCAGGCTCAGATCATCTCCACAGAAAGCCCTCAT 1297
DB 1298 CTACTCTTCTTCTTCTCCAGACAGTGACAAAGGAGCTGTGTGGCATCTGGAC 1357

DB 01-FEB-2002 (first entry)
XX Human breast cell single exon nucleic acid probe #5674.
XX Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
XX Homo sapiens.
XX WO200157271-A2.
XX PD
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX Claim 4; SEQ ID NO 5674; 327pp + sequence listing; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Br 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;
XX Query Match 10.1%; Score 261; DB 22; Length 261;
XX Best Local Similarity 100.0%; Pred. No. 2.2e-64;
XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTGG 1177
DB 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTGG 60
QY 1178 AAGTGAGTGAACACTGCAATGGAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237
DB 61 AAGTGAGTGAACACTGCAATGGAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120
QY 1238 CATCTCTTCAATGTGCTCAGCTCAGGCTCAGATCATCTCCACAGAAAGCCCTCAT 1297
DB 1298 CATCTCTTCAATGTGCTCAGCTCAGGCTCAGATCATCTCCACAGAAAGCCCTCAT 1297
DB 1298 CTACTCTTCTTCTTCTCCAGACAGTGACAAAGGAGCTGTGTGGCATCTGGAC 1357
```

Db 121 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 180
 QY 1298 CTACCTCTTGTCTCTCTCCAGACAGTGAACAAAGCCAGCTCTGTTGAGGCATCTGGAC 1357
 Db 181 CTACTCTCTTGTCTCTCTCCAGACAGTGAACAAAGCCAGCTCTGTTGAGGCATCTGGAC 240
 QY 1358 ACCACACACATCATCATCTG 1378
 Db 241 ACCACACACATCATCATCTG 261

RESULT 23
 ABA31975
 ID ABA31975 standard; DNA; 261 BP.
 XX AC ABA31975;
 XX 23-JAN-2002 (first entry)
 DE Probe #10441 for gene expression analysis in human heart cell sample.
 DE Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX OS Homo sapiens.
 XX PN WO200157274-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00666.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human hearts -
 XX Claim 4; SEQ ID No 10441; 530pp; English.
 XX CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pt_sequences.
 XX SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;
 Query Match 10.1%; Score 261; DB 22; Length 261;
 Best Local Similarity 100.0%; Pred. No. 2.2e-64;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAGTCAATCAGATATTTCACATCTTCTCTAGGTCCTCCATCAAGTCTGATTGG 1177
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 1 GTGGAACAGTCAATCAGATATTTCACATCTTCTCTAGGTCCTCCATCAAGTCTGATTGG 60
 QY 1178 AAGTGAGTGAAGCTGCAATGGAGTAGATACCTCCAGCTGAACAAATTTCTTTCAGCCTTCTA 1237
 Db 61 AAGTGAGTGAAGCTGCAATGGAGTAGATACCTCCAGCTGAACAAATTTCTTTCAGCCTTCTA 120
 QY 1238 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 1297
 Db 121 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 180
 QY 1298 CTACTCTCTTGTCTCTCTCCAGACAGTGAACAAAGCCAGCTCTGTTGAGGCATCTGGAC 1357
 Db 181 CTACTCTCTTGTCTCTCTCCAGACAGTGAACAAAGCCAGCTCTGTTGAGGCATCTGGAC 240
 QY 1358 ACCACACACATCATCATCTG 1378
 Db 241 ACCACACACATCATCATCTG 261

RESULT 24
 AAK13288
 ID AAK13288 standard; DNA; 261 BP.
 XX AC AAK13288;
 XX 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe SEQ ID NO: 13279.
 DE Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 XX OS Homo sapiens.
 XX PN WO200157275-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US00667.
 XX PR 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human brains -
 XX Example 4; SEQ ID NO: 13279; 650pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;
 Query Match 10.1%; Score 261; DB 22; Length 261;
 Best Local Similarity 100.0%; Pred. No. 2.2e-64;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Wed Mar 26 09:38:49 2003

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTTGG 1177
 Db 1 GTGGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTTGG 60
 QY 1178 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTA 1237
 Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTA 120
 QY 1238 CATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGT 1297
 Db 121 CATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGT 180
 QY 1298 CTACTCTTCTTCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTTGAGGCATCTGGAC 1357
 Db 181 CTACTCTTCTTCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTTGAGGCATCTGGAC 240
 QY 1358 ACCACACATCATCAGTCTG 1378
 Db 241 ACCACACATCATCAGTCTG 261

RESULT 25
 AAK39023
 ID AAK39023 standard; DNA; 261 BP.

XX AC AAK39023;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed single exon probe SEQ ID NO: 13580.

XX DE Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX PN WO200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00668.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488900/53.

XX XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -

XX PS Example 4; SEQ ID NO: 13580; 658pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.

XX SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;

Query Match 10.1%; Score 261; DB 22; Length 261;
 Best Local similarity 100.0%; Pred. No. 2.2e-64;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTTGG 1177
 Db 1 GTGGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTGATTTGG 60
 QY 1178 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTA 1237
 Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTA 120
 QY 1238 CATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGT 1297
 Db 121 CATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGT 180
 QY 1298 CTACTCTTCTTCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTTGAGGCATCTGGAC 1357
 Db 181 CTACTCTTCTTCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTTGAGGCATCTGGAC 240
 QY 1358 ACCACACATCATCAGTCTG 1378
 Db 241 ACCACACATCATCAGTCTG 261

RESULT 26

AAI19833
 ID AAI19833 standard; DNA; 261 BP.

XX AC AAI19833;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #9766 for gene expression analysis in human cervical cell sample.
 DE Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488901/53.

XX XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 25; SEQ ID No 9766; 487pp; English.

XX CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;
Query Match 10.1%; Score 261; DB 22; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.2e-64;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1118 GTGGACAAGTCAATCAGATATTTCAACTCTTCCCTACGGTCCCATCAAGTCTGATTGG 1177
DB 1 GTGGACAAGTCAATCAGATATTTCAACTCTTCCCTACGGTCCCATCAAGTCTGATTGG 60
QY 1178 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGGAACAATTTCTTCAGCCTTCTA 1237
DB 61 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGGAACAATTTCTTCAGCCTTCTA 120
QY 1238 CATCTCTACATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAGCCCTCATT 1297
DB 121 CATCTCTACATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAGCCCTCATT 180
QY 1298 CTACTCTCTTCTATCTCTCCACAGAGTGAACAAGGCGAGTCTGTGAGGCATCTGGAC 1357
DB 181 CTACTCTCTTCTATCTCTCCACAGAGTGAACAAGGCGAGTCTGTGAGGCATCTGGAC 240
QY 1358 ACCACACATCATCAGTCTG 1378
DB 241 ACCACACATCATCAGTCTG 261
RESULT 28
AAI45028
ID AAI45028 standard; DNA; 261 BP.
AC AAI45028;
DT 17-OCT-2001 (first entry)
DE Probe #13714 used to measure gene expression in human placenta sample.
DE Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
KW Homo sapiens.
OS WO200157272-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00663.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
DR Human genome-derived single exon nucleic acid probes useful for
DR analyzing gene expression in human placenta.
PT Claim 25; SEQ ID No 13714; 654pp; English.
PS The present invention relates to single exon nucleic acid probes (SENP).
PS The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.


```
Query Match          9.98; Score 255.8; DB 22;
Best Local Similarity 99.3%; pred. No. 9.4e-63; Length 481;
Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

us-09-781-693a-1.rng

Wed Mar 26 09:38:49 2003

CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 SQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;

Query Match 9.9%; Score 255.8; DB 22; Length 481;
 Best Local Similarity 99.2%; Pred. No. 9.4e-63;
 Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1696 ATACACAAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCCAACTCAGGAGAAAGAAAT 1755
 DB 223 ACAGCACAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCCAACTCAGGAGAAAGAAAT 282
 QY 1756 GACCTCAATCTTGATCGCTCTTTGGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 1815
 DB 283 GACCTCAATCTTGATCGCTCTTTGGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 342
 QY 1816 AAGGAACCAAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACACC 1875
 DB 343 AAGGAACCAAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACACC 402
 QY 1876 AATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATATCC 1935
 DB 403 AATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATATCC 462
 QY 1936 ACCAGGACTCTGCTCTTC 1954
 DB 463 ACCAGGACTCTGCTCTTC 481

RESULT 32
 AAK02620
 ID AAK02620 standard; DNA; 481 BP.
 AC AAK02620;
 DT 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe SEQ ID NO: 2611.
 DE Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 OS Homo sapiens.
 OS WO200157275-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00667.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-062366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-483446/52.
 Single exon nucleic acid probes for analyzing gene expression in human
 brains -
 Example 4; SEQ ID NO: 2611; 650pp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 brain. They can be used to measure gene expression in brain cell samples,
 which may enable the diagnosis and improved treatment of nervous system
 diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC

QY 1696 ATACACAAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCCAACTCAGGAGAAAGAAAT 1755
 DB 223 ACAGCACAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCCAACTCAGGAGAAAGAAAT 282
 QY 1756 GACCTCAATCTTGATCGCTCTTTGGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 1815
 DB 283 GACCTCAATCTTGATCGCTCTTTGGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 342
 QY 1816 AAGGAACCAAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACACC 1875
 DB 343 AAGGAACCAAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACACC 402
 QY 1876 AATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATATCC 1935
 DB 403 AATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATATCC 462
 QY 1936 ACCAGGACTCTGCTCTTC 1954
 DB 463 ACCAGGACTCTGCTCTTC 481

RESULT 33
 AAK28054
 ID AAK28054 standard; DNA; 481 BP.
 AC AAK28054;
 DT 06-NOV-2001 (first entry)
 DE Human bone marrow expressed single exon probe SEQ ID NO: 2611.
 DE Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 OS Homo sapiens.
 OS WO200157276-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00668.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-062366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488900/53.
 Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human bone marrow -
 Example 4; SEQ ID NO: 2611; 658pp + Sequence Listing; English.
 The present invention provides a number of single exon nucleic acid
 probes which are derived from genomic sequences expressed in the human
 bone marrow. They can be used to measure gene expression in bone marrow
 CC

CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

Query Match 9.9%; Score 255.8; DB 22; Length 481;
Best Local Similarity 99.2%; Pred. No. 9.4e-63;
Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

223	ACAGCAAAATCAGATAAGTTCCACAGCCAAAGCCATTGGATTC	1755
223	ACAGCAAAATCAGATAAGTTCCACAGCCAAAGCCATTGGATTC	282
1756	GACCTCAATCTTGTATCGCTCTTTGTGGGTTCCAGAGAATCTGCTTCATCTGAAAAAGCC	1815
283	GACCTCAATCTTGTATCGCTCTTTGTGGGTTCCAGAGAATCTGCTTCATCTGAAAAAGCC	342
1816	AAGGAACCCAGAAACTTTCAGATCAGACTAGCAGTGTACCAATGAAATAACACCC	1875
343	AAGGAACCCAGAAACTTTCAGATCAGACTAGCAGTGTACCAATGAAATAACACCC	402
1876	AATCTGTAGCCTCAGTTTCCAAACAGAGCCACTGGGCTTCAGTCTATGAAGAATATCC	1935
403	AATCTGTAGCCTCAGTTTCCAAACAGAGCCACTGGGCTTCAGTCTATGAAGAATATCC	462
1936	ACGAGGAGACTGTGCTTTC	1954
463	ACGAGGAGACTGTGCTTTC	481

SULT 34
I12631
AAI12631 standard; DNA; 481 BP.

AA112631;
12-OCT-2001 (first entry)

Probe #2564 for gene expression analysis in human cervical cell sample.
Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer; ss.

Homo sapiens.
WO200157278-A2.

09-AUG-2001. 30-JAN-2001: 2001WO-HSC00570

04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0207456.
01-JUL-2000; 2000US-0207456.

03-AUG-2000; 2000US-0608408.
21-SEP-2000; 2000US-0632366.
27-SEP-2000; 2000US-0234687.
04-OCT-2000; 2000US-0236359.

01 OCT 2000; 2000GB-0024263.
(MOLE-) MOLECULAR DYNAMICS INC.
DORA 50

Fenn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID No 256A; 487221

The present invention relates to human single exon nucleic acid probes (SENP). The present sequence is one such probe. The SENPs are defined

...ents are derived

from human HeLa cells. The SENSEs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from ftp.wipo.int/pub/published_pat_sequences.

ISQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;
Query Match
Best Local Similarity 9.9%; Score 255.8; DB 22; Length 40

Local Similarity 99.2%; Pred. No. 9.4e-63;
 Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0

b
y

223 ACAGCACAAATCAGATAAGTTTCACAGCCAGCCATTGGATTCCAATCAGGAGAAAAGAAT 1755
1756 GACCCTCAATCTTGATCGGTCTTCTGGGGTTCACAGAACACTCATGCCTCTCTGAATTAAT 282

283 GACCTCAATCTTGATCGCTCTTCTGGGGTCCAGAAGAACTGCTTCATCTGAAAAAGCC 342

343 AAGGAACCGAAGAACTTCAGATCAGCTAGCCACTGAGAGTGCTACCAATGAAATATACACC 1875

1876 AATCTCTGAGCCTCAGTCTCCAGACGCTGAGAGTGCTACCAATGAAATATACACC 402

|||||CCCCCAGAACCCACCTGGGCCCTTCAGTCTCATGAAGAATACC 1935
403 AATCCTGAGCCTCAGTTCCAAACAGAACCCACTGGGCCCTTCAGTCTCATGAAGAATACC |||||
1936 ACCAGGGGACTGTGCATCGAACA 1937

463 ACCAGGGACTCTGCTCTTC 481

SULT 35
I02537
AAI02537 standard; DNA; 481 bp.

AAI02537;
09-OCT-2001 (first entry)

Probe #2528 used to measure gene expression in human breast sample.
Probe; human; breast disease; breast cancer

inflammatory disease; proliferative breast disease; non-carcinoma tumour.
Homo sapiens.

WO200157270-A2.
09-AUG-2001.

29-JAN-2001; 2001WO-US00661.
04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0180312.

30-JUN-2000; 2000US-0207456.
03-AUG-2000; 2000US-0608408.
21-SEP-2000; 2000US-0632366.
27-SEP-2000; 2000US-0234687.

(MOLE-) MOLECULAR DYNAMICS INC.
27 SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression in a human breast -

XX	Homo sapiens.
OS	WO200186003-A2.
XX	
XX	
PN	
XX	
XX	
PD	15-NOV-2001.
XX	
XX	
PF	30-JAN-2001; 2001WO-US000655.
XX	
PR	04-FEB-2000; 2000US-180312P.
XX	

CC	XX	SQ	Sequence	481 BP;	152 A;	106 C;	89 G;	134 T;	0 other;
			Query Match	9.98;	Score	255.8;	DB	24;	Length 481;
			Best Local Similarity	99.2%;	Pred.	No. 9.4e-63;			
			Matches	257;	Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
QY		1696	ATACACAATCAGATAAGTTCACAGCCCAAGCCATTGGATTCCCAACTCAGAGAGAAAGAAAT	1755					
Db		223	ACAGCACATCAGATAAGTTCACAGCCCAAGCCATTGGATTCCCAACTCAGAGAGAAAGAAAT	282					
QY		1756	GACCTCAACTTGTATCGCTCTTGTGGGTTCCAGAAAGAAATCTGCTCATCTGAAAAGCC	1815					
Db		283	GACCTCAACTTGTATCGCTCTTGTGGGTTCCAGAAAGAAATCTGCTCATCTGAAAAGCC	342					

QY 1816 AAGGAACCGAAGAACTTCTAGATCAGACTAGCAGTCTACCAATGAAATTAACACC 1875
 Db 343 AAGGAACCGAAGAACTTCTAGATCAGACTAGCAGTCTACCAATGAAATTAACACC 402
 QY 1876 AATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCTTTCAGCTCATGAGAAACATCC 1935
 Db 403 AATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCTTTCAGCTCATGAGAAACATCC 462
 QY 1936 ACCAGGACTCTGCTCTTC 1954
 Db 463 ACCAGGACTCTGCTCTTC 481

RESULT 37
 ABV60730

ID ABV60730 standard; cDNA; 423 BP.

AC ABV60730;

DT 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 60721.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW Pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US05171.

PR 17-FEB-2000; 2000US-183319P.

PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.

PR 03-JUN-2000; 2000US-211314P.

PR 18-JUL-2000; 2000US-219007P.

PR 13-DEC-2000; 2000US-255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer

PS Claim 1; Page 11554-11555; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) determining whether prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;

CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 423 BP; 152 A; 87 C; 93 G; 91 T; 0 other;

SQ Query Match 9.68; Score 248.4; DB 23; Length 423;

Best Local Similarity 97.7%; Pred. No. 1.2e-60;

Matches 252; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1366 CATCATCAGTCTGATACAAATATGAAAGCTGAGCCCAACCCAGGACAGGTGAACCA 1425
 Db 166 CATATCCGAGCTGATACAAATATGAAAGCTGAGCCCAACCCAGGACAGGTGAACCA 225
 QY 1426 GTTTTAAGTTTGCACCTACAGCAGAGAGGAACTACAAAGCAGCAATAAAACTGAACCTT 1485
 Db 226 GTTTTAAGTTTGCACCTACAGCAGAGAGGAACTACAAAGCAGCAATAAAACTGAACCTT 285
 QY 1486 ACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTGGAGCCATTGCAATCT 1545
 Db 286 ACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTGGAGCCATTGCAATCT 345
 QY 1546 GAGGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTGCACCAAGAGGAGACAGT 1605
 Db 346 GAGGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTGCACCAAGAGGAGACAGT 405
 QY 1606 GAAACAAAAGCTCTCGAA 1623
 Db 406 GAAACAAAAGCTCTCGAA 423

RESULT 38
 ABA46980

ID ABA46980 standard; DNA; 186 BP.

AC ABA46980;

XX 01-FEB-2002 (first entry)

DE Human breast cell single exon nucleic acid probe #5675.

KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes

PS Claim 4; SEQ ID NO 5675; 327pp + sequence listing; English.

CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Br 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosis breast disease. Gene

Wed Mar 26 09:38:49 2003

measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a single exon nucleic acid probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;

Query Match 7.2%; Score 186; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.8e-43;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTTGACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAGAAAGTTCTTACT 312
DB 1 GTTTTGACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAGAAAGTTCTTACT 60
QY 313 TGTACAAATGATAAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 372
DB 61 TGTACAAATGATAAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 120
QY 373 GTTGAGCAAGATGCAGAAACCAACAGACATGCCAATTTACGTGTCATTATGGAATCT 432
DB 121 GTTGAGCAAGATGCAGAAACCAACAGACATGCCAATTTACGTGTCATTATGGAATCT 180
QY 433 TATGAG 438
DB 181 TATGAG 186

RESULT 40
ABA31976
ID ABA31976 standard; DNA; 186 BP.

XX ABA31976;
XX 23-JAN-2002 (first entry)
DE Probe #10442 for gene expression analysis in human heart cell sample.
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX Homo sapiens.
OS WO200157274-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00666.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488899/53.

Single exon nucleic acid probes for analyzing gene expression in human hearts -
Claim 4; SEQ ID No 10442; 530pp; English.

expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequences. The present sequence is a single exon nucleic acid probe of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;
Query Match 7.2%; Score 186; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 5.8e-43;
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTTGACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAGAAAGTTCTTACT 312
DB 1 GTTTTGACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAGAAAGTTCTTACT 60
QY 313 TGTACAAATGATAAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 372
DB 61 TGTACAAATGATAAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 120
QY 373 GTTGAGCAAGATGCAGAAACCAACAGACATGCCAATTTACGTGTCATTATGGAATCT 432
DB 121 GTTGAGCAAGATGCAGAAACCAACAGACATGCCAATTTACGTGTCATTATGGAATCT 180
QY 433 TATGAG 438
DB 181 TATGAG 186

RESULT 39
ABA64861
ID ABA64861 standard; DNA; 186 BP.

XX ABA64861;
XX 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #13166.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX Homo sapiens.
OS WO200157277-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US00669.
PF 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver -
Claim 4; SEQ ID NO 13166; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for

11

12

13

14

15

16

17

18

19

20

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2003, 23:21:03 ; Search time 4054 seconds
(without alignments)
10306.945 Million cell updates/sec

Title: US-09-781-693a-1

Perfect score: 2580

Sequence: 1 atgtctcgggggtgctccta.....atgaataatgagatgagaa 2580

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1981.4	76.8	2375	11 AK004618	AK004618 Mus muscu
2	890.8	34.5	1045	14 BM807192	BM807192 AGENCOURT
3	853.6	33.1	1052	13 BM477718	BM477718 AGENCOURT
4	832.6	32.3	864	14 BQ433226	BQ433226 AGENCOURT
5	827.4	32.1	853	14 BQ933148	BQ933148 AGENCOURT
6	826.6	32.0	1074	14 BQ067602	BQ067602 AGENCOURT

7	784.8	30.4	839	12	BG740812	BG740812
8	781.4	30.3	805	12	BG686945	BG686945
9	743	28.8	827	13	BI335444	BI335444
10	735.8	28.5	951	12	BF980168	BF980168
11	725.8	28.1	843	14	BQ229691	BQ229691
12	723	28.0	906	14	BQ433486	BQ433486
13	720.4	27.9	722	9	AL560929	AL560929
14	716.8	27.8	791	13	BI834692	BI834692
15	701.2	27.2	860	13	BI462541	BI462541
16	684.2	26.5	809	13	BI464753	BI464753
17	681.4	26.4	730	13	BI461216	BI461216
18	678.8	26.3	729	12	BG682657	BG682657
19	677.4	26.3	783	13	BI464045	BI464045
20	667.8	25.9	754	12	BG718960	BG718960
21	665.8	25.8	834	13	BI461627	BI461627
22	663.6	25.8	850	13	BI464622	BI464622
23	658.2	25.5	694	10	BE180558	BE180558
24	655.8	25.4	739	12	BG117357	BG117357
25	648.6	25.1	1075	14	BM923018	BM923018
26	646.6	25.1	775	12	BE867206	BE867206
27	641.8	24.9	806	12	BG772319	BG772319
28	641.6	24.9	911	12	BG530857	BG530857
29	640	24.8	648	14	BQ019842	BQ019842
30	634	24.6	634	12	BE747286	BE747286
31	633.4	24.6	635	14	BM790042	BM790042
32	631.4	24.5	707	12	BG719267	BG719267
33	629.8	24.4	728	12	BG534181	BG534181
34	628.6	24.4	869	13	BI459472	BI459472
35	627.4	24.3	762	10	BE559905	BE559905
36	627	24.3	717	13	BI559857	BI559857
37	624.4	24.2	733	13	BI561174	BI561174
38	624	24.2	1199	14	BM926145	BM926145
39	623.6	24.2	666	10	BE180608	BE180608
40	618.4	24.0	912	12	BF029089	BF029089
41	616.6	23.9	709	12	BG720740	BG720740
42	612.8	23.8	868	12	BE884542	BE884542
43	611.6	23.7	918	12	BI65887	BI65887
44	610.8	23.7	616	13	BI836271	BI836271
45	605	23.4	624	10	BE180560	BE180560
46	601.6	23.4	604	9	AL602747	AL602747
47	601.6	23.3	735	12	BF672818	BF672818
48	599.2	23.2	718	13	BI562377	BI562377
49	596.8	23.1	730	12	BG724412	BG724412
50	592.4	23.0	847	12	BI64080	BI64080
51	588	22.8	597	14	BM726957	BM726957
52	584.4	22.7	1141	12	BG255311	BG255311
53	582.8	22.6	620	14	BQ575260	BQ575260
54	580.6	22.5	922	12	BE882087	BE882087
55	575.4	22.3	668	13	BM538035	BM538035
56	574.8	22.3	593	9	AI832097	AI832097
57	574	22.2	619	12	BF695020	BF695020
58	573.2	22.2	676	12	BF979546	BF979546
59	567.8	22.0	791	12	BG484572	BG484572
60	566.6	22.0	956	12	BF662916	BF662916
61	562.4	21.8	658	10	BE268392	BE268392
62	557.4	21.6	578	10	BE180559	BE180559
63	550	21.3	805	12	BG217514	BG217514
64	547	21.2	547	9	AI927915	AI927915
65	535	20.7	713	10	BE545703	BE545703
66	511.6	19.8	550	10	AW992096	AW992096
67	510.6	19.8	532	10	BE019647	BE019647
68	509	19.7	832	14	BQ435888	BQ435888
69	508	19.7	508	10	BE504075	BE504075
70	507.6	19.7	839	14	BQ221011	BQ221011
71	504.2	19.5	540	10	AW992159	AW992159
72	504.2	19.5	675	12	BF671543	BF671543
73	504	19.5	504	9	AA626631	AA626631
74	488.2	19.3	1132	13	BM478778	BM478778
75	486.2	19.2	567	9	AUI46956	AUI46956
76	492.8	19.1	544	9	AUI60319	AUI60319
77	490.8	19.0	510	10	BE019684	BE019684
78	488	18.9	516	14	BM685550	BM685550
79	487.8	18.9	711	12	BG567316	BG567316

wed Mar 26 09:38:52 2003

MEDLINE
PUBMED
REFERENCE
AUTHORS

20499374
11042159
3

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiyosawa, H., Adachi, J., Fukuda, S., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujimori, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Adachi, J., Fukuda, S., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Quackenbush, J., Schriml, L., M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Hill, D., Hofmann, C., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Marzari, J., Mombarts, P., Sasaki, H., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Storch, K.F., Suzuki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

5

(bases 1 to 2975)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hara, A., Harada, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamashita, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Tel: 81-45-503-9222, Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

1

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

COMMENT

1

ALIGNMENTS

RESULT 1
AK004618
LOCUS
DEFINITION
Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:120006M05; homolog to DJ703H14.1 PROTEIN (FRAGMENT), full insert sequence.
2975 bp mRNA linear HTC 19-JAN-2002

AK004618
AK004618.1 GI:12835911
Version
Keywords
Source
Mus musculus (strain: C57BL/6J) adult male lung cDNA to mRNA, HTC; CAP trapper.
Mus musculus (strain: C57BL/6J) adult male lung cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library
clone:120006M05.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

80 485.4 18.8 559 14 W19888
81 485.4 18.8 823 9 AU118130
82 483 18.7 844 10 BE568492
83 480.4 18.6 510 9 AL705203
84 473 18.3 548 12 BG101295
85 473 18.3 575 9 AU144811
86 467.8 18.1 662 9 AU705761
87 467.4 18.0 469 10 AW087744
88 465 18.0 524 10 BE397963
89 461.2 17.9 550 10 AW741120
90 461.2 17.9 713 12 BE671187
91 457.4 17.7 757 13 BL155467
92 451.8 17.5 855 12 BF575729
93 446.4 17.2 448 12 BF062454
94 444.6 17.1 548 12 BG719120
95 441 17.1 516 13 BM539925
96 438.4 17.0 454 9 AA47852
97 437 16.9 454 9 AA47852
98 436.8 16.9 834 12 BG775713
99 434.2 16.8 570 9 AA479513
100 431.4 16.7 480 12 BF354864
101 431.4 16.7 778 12 BG750040
102 430.8 16.7 443 10 BE041839
103 430.6 16.7 437 10 BE175440
104 430 16.7 505 14 W00613
105 424.4 16.4 438 14 N54388
106 421.8 16.3 729 10 BE332778
107 421 16.3 469 12 BE718001
108 420.4 16.3 448 10 AW119185
109 413.8 16.0 425 10 BE175441
110 413.8 16.0 454 9 A1334344
111 413.8 16.0 744 13 BI464683
112 405.4 15.7 646 10 BB617380
113 393.8 15.3 420 13 BM663460
114 401 15.5 726 12 BG539579
115 393.8 15.3 422 9 AA085991
116 392 15.2 414 10 AW771580
117 391.6 15.2 527 9 AL568781
118 390.2 15.1 506 13 BM538697
119 388.8 15.1 460 12 BF561661
120 388.4 15.1 416 9 AA447700

||||| 2659 CGAGTCACCGTCTGGAGGCTGACAGATCAGAGGTTTCAGGTCAGGAGATGAAATGAG 2718
Db 2659 GATGAGGAA 2580
QY 2572 GATGAGGAA 2580
Db 2719 GATGAGGAA 2727
RESULT 2
BM807192 1045 bp mRNA linear EST 05-MAR-2002
AGENCOURT_6552622 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5552090
5', mRNA sequence.
BM807192
BM807192.1 GI:19124008
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1045)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cchapman@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12266 Row: O Column: 03
High quality sequence stop: 693.
location/Qualifiers
1. 1045
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5552090"
/clone.lib="NIH_MGC_85"
/tissue.type="lymphoma, cell line"
/lab.host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-df primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library." 1 others
BASE COUNT 331 a 222 c 251 g 240 t
ORIGIN
Query Match 34.5%; Score 890.8; DB 14; Length 1045;
Best Local Similarity 96.0%; Pred. No. 5.5e-241;
Matches 978; Conservative 0; Mismatches 33; Indels 8; Gaps 6;
QY 640 AGCTCAGTACGAAATATATGATCGCGGAATCTGGGACAGAGCTACAGGAAATATGCA 699
Db 1 AGCTCAGTACGAAATATATGATCGCGGAATCTGGGACAGAGCTACAGGAAATATGCA 60
QY 700 GGTCTCAGGAGTACTGGAAATGTTGCCCGTTTATCTCCCATCTTATATAAGTCC 759
Db 61 GGTCTCAGGAGTACTGGAAATGTTGCCCGTTTATCTCCCATCTTATATAAGTCC 120
QY 760 TGCAGAGTGCATCTCTGTGTACAGTGAAGTGTCAAGAGATCTCGTTAGTTACTCT 819
Db 121 TGCAGAGTGCATCTCTGTGTACAGTGAAGTGTCAAGAGATCTCGTTAGTTACTCT 180
QY 820 TCAGATTACATATCTTTTTCACCCGAAAGATGATACAGCAGAGCTTAAACCTCT 879
Db 181 TCAGATTACATATCTTTTTCACCCGAAAGATGATACAGCAGAGCTTAAACCTCT 240
QY 880 TCTCCGAGAGAGAGAGAGAGAGTTCGCCACACACAGTAAAGCGTTTGAAGCTTCGT 939
|||||

1501 AGTATPAGCATCAAGTTCTAGAGGATTTGGAGGCAATTCGCAAAATCTGAGGCTCAGGAGGAA 1560
Db 1597 AGTACAGCTCAAGTTCCAGAGGAAATTTGGAGGCAATTCGCAAAATCTGAGGCTCAGGAGGAA 1656
QY 1561 TCTTTCTGCCACAGAGCTCAGTGCACCCAGAGGAGAGAGTGAACAAAAGCTCTCT 1620
Db 1657 TCGTTGGTCCCTCCGAGCTCTGTGCGCCACCGGAGAGAGAGTGAACAAAAGCTCTCT 1716
QY 1621 GAAGAATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAACCCAGTT 1680
Db 1717 GAAGAATCATCAG- 1758
QY 1681 GAGAACCATATCAATATACACAAATCAGATAGTTCACAGCCAGCCATTGGATTCCAAC 1740
Db 1759 CAATACAGATATACAGCACAACTTGATACTTCCCGCTGAGCCATTGGATTCTAAC 1818
QY 1741 TCAGGAGAGAAATGACCTCAATCTTGATCGCTCTGTGGGGTTCAGAGAAATCTGCT 1800
Db 1819 TCAGGAGAGAAATGACCTCAATCTTGATCGCTCTGTGGGGTTCAGAGAAATCTGCT 1878
QY 1801 TCATCTGAAAAGCAAGGAAAC 1822
Db 1879 TTGCTGAACAGACAGAGGAGACTTGTGAGCGCCAGACACTGAGTGTACCGGCAAT 1938
QY 1823 -----CAGAACTTCAGATCAGACTGAGTGTACCGGCAAT 1851
Db 1939 GGTAGCACAGCGCTGAATCCCATCCAGACAGAGAGGACTTGTGAGCGCCAGACACTGAG 1998
QY 1852 AGTGTACCAATGAAATTAACACCAATCTGCTGAGTTCACCAAGAGGAGGAGGAGG 1911
Db 1999 AGTGTACCAAGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2058
QY 1912 CTTTCACTCATGAGAAACATCCACAGGAGTCTGCTTTCAGGACAGATGACAGT 1971
Db 2059 CTTTCACTCATGAGAAACATCCACAGGAGTCTGCTTTCAGGACAGATGACAGT 2118
QY 1972 GATGATGACCCAGTCTGTATCCAGGTCAGAGTATCCAGGAGGAGGAGGAGGAGGAGG 2031
Db 2119 GATGATGATCCGCTCTGTATCCAGGTCAGAGTATCCAGGAGGAGGAGGAGGAGGAGG 2178
QY 2032 TCTGCTGTTCCCGTATTCAGGAGTCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2091
Db 2179 TCCGCTGTTCCCGATTCAGGAGTCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2238
QY 2092 GAATTTGATATCTTGAACATTAAGAGGCGCTAGTAAATGTTTATTAAGGCGCATCGC 2151
Db 2239 GAGCTGGATATCTTTGAACATTAAGAGGCGCTAGTAAATGTTTATTAAGGCGCATCGC 2298
QY 2152 AACTCCAGGAGCAATGATAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2211
Db 2299 AACTCCAGGAGCAATGATAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2358
QY 2212 TCTGATGTGGCCACATTTTCATCTGCGATCGGACACTGCTGAGCAATTTGATGCTCTG 2271
Db 2359 TCCGATTTGGCCATATCTTCATCTGCGATCGGACACTGCTGAGCAATTTGATGCTCTG 2418
QY 2272 GAAGCTGATAATCATGTGTGAATGCTGAGCCACATCCGTTTGAACCAATTTAGCC 2331
Db 2419 GAAGCTGATAATCATGTGTGAATGCTGAGCCACATCCGTTTGAACCAATTTAGCC 2478
QY 2332 TCATCTGGCATGATATGACATAAGATCTGTCACCATTTAGAGAGTCAAGGATTTT 2391
Db 2479 TCATCTGGCATGATATGACATAAGATCTGTCACCATTTAGAGAGTCAAGGATTTT 2538
QY 2392 AACCGAAACTTCTGCTGATGAAGTTATTAACTCGAACGACATCTGCTGGAAGAACTAGA 2451
Db 2539 AATCGAAACTTCTGCTGATGAAGTTATTAACTCGAACGACATCTGCTGGAAGAGCTCGG 2598
QY 2452 AACACCATACAGTCCAGCTCTTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATC 2511
Db 2599 AACACCATACAGCTCCAGCTCTTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATC 2658
QY 2512 CGAGCTGACCGGTTTGGAGGCTGACAGATCAGAGGCTCTGGTCAAGAGAAATGAAATGAG 2571
|||||

Wed Mar 26 09:38:52 2003

```

Db 721 CAGATGAATGGAGCAGTAGCATCAAGTTCTAGAGGAATTTGGAGGCAATTCGCAAAATCTG 780
QY 1547 AGGTCAGGAGGAATTTTCGTCCTCCACAGAGCTCAGTCAACCA-CCAGAAGGAGACAGT 1605
Db 781 AGGTCAGGAGGAATTTTCGTCCTCCACAGAGCTCAGTCAACCA-CCAGAAGGAGACAGT 840
QY 1606 GAAACAAAAGCTCCTCAAGAAATCATCAGAGGATGTG---ACAAAATATCAGGAAGGATGA 1662
Db 841 GAAACAAAAGCTCCTCAAGAAATCATCAGAGGATGTGCAACAAATATTCAGGAAGGATGA 900
QY 1663 TC-TGCAGAAAACCCAGTTGAGCA---ACCATATCAATATAACCAATC 1706
Db 901 TCTTGCAGAAAACCCAGTTGAGAGCCCTTATCCATATAACCAATC 948

RESULT 4
BQ433226 864 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7774746 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6023741
DEFINITION 5', mRNA sequence.
ACCESSION BQ433226
VERSION BQ433226.1 GI:21172302
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 864)
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCPD/DPG/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13233 row: c column: 06
High quality sequence stop: 692.
Location/Qualifiers
1. .864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6023741"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 272 a 188 c 206 t
ORIGIN

Query Match 32.3%; Score 832.6; DB 14; Length 864;
Best Local Similarity 99.3%; Pred. No. 1.6e-224;
Matches 857; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1686 CCATATCAATATACATCAGATAGTTCCACAGCAAGCCATTGGATTCCCAACTCAGG 1745
Db 1 CCATATCAATATACATCAGATAGTTCCACAGCAAGCCATTGGATTCCCAACTCAGG 60
QY 1746 AGAAGAAATGACCTCAATCTTGTATCGCTCTGTGGGGTTCACAGAAATCTGCTTCATC 1805
Db 61 AGAAGAAATGACCTCAATCTTGTATCGCTCTGTGGGGTTCACAGAAATCTGCTTCATC 120
QY 1806 TGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGAGTGCTACCAATGA 1865
Db 121 TGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGAGTGCTACCAATGA 180

```

Query Match	32.1%	Score 827.4	DB 14	Length 853
Best Local Similarity	98.7%	Pred. No. 4.8e-223		
Matches 842	Conservative 0	Mismatches 10	Indels 1	Gaps 1
QY 1249	ATGTACGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTTACTCCTTTG	1308		
DB 1	ATGTACGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTTACTCCTTTG	60		
QY 1309	CTATCTTCTCAGACAGTGAACAAAGCAGTCTGTGTGAGGCATCTGGACACACACACAT	1368		
DB 61	CTATCTTCTCAGATAGTGAACAAAGCAGTCTGTGTGAGGCATCTGGACACACACACAT	120		
QY 1369	CATCAGTCTGATACAAATATGAAAGCTCAGGCCCAACCCAGGACAGGTGAACCAAGTT	1428		
DB 121	CATCAGTCTGATACAAATATGAAAGCTCAGGCCCAACCCAGGACAGGTGAACCAAGTT	180		
QY 1429	TTAAGTTTGCACCTACAGCAGAGGAACCACTACAAGCACAAATAAAACTGAACTTTACA	1488		
DB 181	TTAAGTTTGCACCTACAGCAGAGGAACCACTACAAGCACAAATAAAACTGAACTTTACA	240		
QY 1489	GATTGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTCGAATCTGAG	1548		
DB 241	GATTGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTCGAATCTGAG	300		
QY 1549	GGTCAGAGGAATCTTTTCGTCACAGAGCTCAGTGCACCAACCCAGAAAGGAGACAGTGAA	1608		
DB 301	GGTCAGAGGAATCTTTTCGTCACAGAGCTCAGTGCACCAACCCAGAAAGGAGACAGTGAA	360		
QY 1609	ACAAAAGCTCTCTCAAGAATCATCAGAGGATGTACAAAATATCAGGAAGGAGATCTGCA	1668		
DB 361	ACAAAAGCTCTCTCAAGAATCATCAGAGGATGTACAAAATATCAGGAAGGAGATCTGCA	420		
QY 1669	GAAAACCCAGTTGAGAACCATATCAATATACAACTCAGATTAAGTTTCAGAGCCAAAGCCA	1728		
DB 421	GAAAACCCAGTTGAGAACCATATCAATATACAACTCAGATTAAGTTTCAGAGCCAAAGCCA	480		
QY 1729	TTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCA	1788		
DB 481	TTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCA	540		
QY 1789	GAAGAATCTGCTTCATCTGAAAAAGCCAGGAACCAAGAACTTCAGATCAGCTTAGCACT	1848		
DB 541	GAAGAATCTGCTTCATCTGAAAAAGCCAGGAACCAAGAACTTCAGATCAGCTTAGCACT	600		
QY 1849	GAGATGCTACCAATGAAATACACCAATCTTGAGCCTCAGTTTCCAAACAGAGCACT	1908		
DB 601	GAGATGCTACCAATGAAATACACCAATCTTGAGCCTCAGTTTCCAAACAGAGCACT	660		

```
QY 1229 AGCCTTCTACATCTCTACATATGTCAGCTCAGCTCATTCGACATCATCTCCACAGAAA 1288
Db 80 AGCCTTCTACATCTCTACATATGTCAGCTCAGCTCATTCGACATCATCTCCACAGAAA 139
QY 1289 GCGCTCATCTACTCTCTTGGCTATCTCTCCAGACAGTGAACAAAGCGAGTCTGTGAGG 1348
Db 140 GCGCTCATCTACTCTCTTGGCTATCTCTCCAGACAGTGAACAAAGCGAGTCTGTGAGG 199
QY 1349 CATCTGGACACACACACATCATCTGATGATGATGATGATGATGATGATGATGATGATG 1408
Db 200 CATCTGGACACACACACATCATCTGATGATGATGATGATGATGATGATGATGATGATG 259
QY 1409 CAGGACAGGTGAACAGGTTTAAAGTTTGCATCTACGACAGAGGAACAACTACAGCA 1468
Db 260 CAGGACAGGTGAACAGGTTTAAAGTTTGCATCTACGACAGAGGAACAACTACAGCA 319
QY 1469 CAATAAACTGAACCTTTACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATG 1528
Db 320 CAATAAACTGAACCTTTACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATG 379
QY 1529 GGAGCCATTGCAAACTCTGAGGGTCAGGAGGAATCTTTCGTCACACAGAGCTCAGTGAAC 1588
Db 380 GGAGCCATTGCAAACTCTGAGGGTCAGGAGGAATCTTTCGTCACACAGAGCTCAGTGAAC 439
QY 1589 CACCAGAGGAGACAGTGAACAAAGCTCTCTGAAGATCATCAGAGGATGTGACAAAAT 1648
Db 440 CACCAGAGGAGACAGTGAACAAAGCTCTCTGAAGATCATCAGAGGATGTGACAAAAT 499
QY 1649 ATCAGAGAGGATCTCGAGAAACCCAGTTCAGAACCATATCAATATACACATCAG 1708
Db 500 ATCAGAGAGGATCTCGAGAAACCCAGTTCAGAACCATATCAATATACACATCAG 559
QY 1709 ATAAGTTTCAGAGCAAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTG 1768
Db 560 ATAAGTTTCAGAGCAAGCCATTGGATTCCAACTCANGAGAAAGAAATGACCTCAATCTTG 619
QY 1769 ATGCTCTTGTGGGTTCCAGAAATCTGCTTCATCTGAAAAGCAA--GGAACAGAA 1827
Db 620 ATGCTCTTGTGGGTTCCAGAAATCTGCTTCATCTGAAAAGCAAAGGAAACAGAA 679
QY 1828 ACTTCAGATCAGACTAGCAGTGTGCTTACCAATGAAATGAAATGAAATGAAATGAAATG 1887
Db 680 ACTTCAGATCAGACTAGCAGTGTGCTTACCAATGAAATGAAATGAAATGAAATGAAATG 739
QY 1888 CAGTTCCAAACAGAGCACTGGGCTTACGCTCAGTCAATGAAAGAAATCCACAGGAGTCT 1947
Db 740 CAGTTCCAAACAGAGCACTGGGCTTACGCTCAGTCAATGAAAGAAATCCACAGGAGTCT 799
QY 1948 GCTCTTCAGACACAGTACAGTGTGATGATGATGATGATGATGATGATGATGATGATG 2006
Db 800 GCTCTTCAGACACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859
QY 2007 TCGAGCAGGACCTGGTATAGAGCTCTGCT-GTTGCCCGTATTTCAGGAGTCTTCAGAC 2065
Db 860 TCAACCAAGACCTGGTGAATAGCTCTGCTGTTGCCCGTATTTCGGGAATCTTCCTCAA 919
QY 2066 GGAGAAAGAAAGAAAGAAATGGA 2091
Db 920 ACGGGAGAAAAAGAAAGAAAAA 945
```

```
RESULT 7
BG740812 839 bp mRNA linear EST 15-MAY-2001
LOCUS 602632794F1 NCI_CGAP_Skn3 Homo sapiens cdna clone IMAGE:477894 5',
DEFINITION mRNA sequence.
ACCESSION BG740812
VERSION BG740812.1 GI:14051465
KEYWORDS human.
SOURCE EST.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE

1 (bases 1 to 839)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10632 row: 1 column: 23
High quality sequence stop: 819.

FEATURES

source
1..839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4777894"
/lab_host="NCI_CGAP_Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dr.
Average insert size 1.5Kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 265 a 181 c 193 g 200 t
ORIGIN

Query Match 30.4%; Score 784.8; DB 12; Length 839;
Best Local Similarity 98.3%; Pred. No. 5.7e-211;
Matches 814; Conservative 0; Mismatches 11; Indels 2; Gaps 2;
QY 659 ATCGCGGAATGCTGGGCACAAAGAGCTACAGGGAATATGACAGTCGAGGACTACTGAA 718
Db 1 ATCGCGGAATGCTGGGCACAAAGAGCTACAGGGAATATGACAGTCGAGGACTACTGAA 60
QY 719 TGGTGGCCGTTTATTCCTCCATCTTAATAAGTCTCTGAGAGTGCATCTCTGT 778
Db 61 TGGTGGCCGTTTATTCCTCCATCTTAATAAGTCTCTGAGAGTGCATCTCTGT 120
QY 779 GTTACAGTGAAGATGTTCAA-GAGATTCTCTAGTTACTTTCAGATTACATATATCTT 837
Db 121 GTTACAGTGAAGATGTTCAAATGATGATTCCTGTAAGTTTCTTCAGATTACATATCTT 180
QY 838 TTTGACCCGAAAGATGATACAGCAGCAAGAACTTAAACTCTCTTCGCGAAGAGAGA 897
Db 181 TTTGACCCGAAAGATGATACAGCAGCAAGAACTTAAACTCTCTTCGCGAAGAGAGA 240
QY 898 GAAGAGTTGCGAACACCAGTTAAGCTTTGAGACTTCGTGGTGGTGGTGGTGGTGGTGGT 957
Db 241 GAAGAGTTGCGAACACCAGTTAAGCTTTGAGACTTCGTGGTGGTGGTGGTGGTGGTGGT 300
QY 958 GGACCCAGAGCAAGCGGAGAGTGAACGAGAGAGATGGAGCAGAGTCCCAATGTG 1017
Db 301 GGACCCAGAGCAAGCGGAGAGTGAACGAGAGAGATGGAGCAGAGTCCCAATGTG 360
QY 1018 TCATTGATGACAGAGATCTCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGAGTT 1077
Db 361 TCATTGATGACAGAGATCTCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGAGTT 420
QY 1078 GCACAAAGCAATAGAGCAGGAGAGATCTGACCCAGAGTGGAAACAGTCAATCAGAT 1137
Db 421 GCACAAAGCAATAGAGCAGGAGAGATCTGACCCAGAGTGGAAACAGTCAATCAGAT 480
QY 1138 ATTTCAACTCTTCTACGGTCCCAATCAAGTCTCTGATTGGAAGTGAAGTGAAGTGAAGT 1197
Db 481 ATTTCAACTCTTCTACGGTCCCAATCAAGTCTCTGATTGGAAGTGAAGTGAAGTGAAGT 540
QY 1198 GAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTTACATCTCTACAAATGTGAGCT 1257
Db 541 GAAGTAGATATCTCCAGCTGAACAAATTTCTTACGCTTCTTACATCTCTTACAAATGTGAGCT 600

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM11346 row: c column: 09
 High quality sequence stop: 754.
 Location/Qualifiers

FEATURES

source
 1. .827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5140544"
 /clone_lib="NIH_MGC_12"
 /tissue_type="cervical carcinoma cell line"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.4 kb. Library prepared by Life
 Technologies."
 BASE COUNT 253 a 160 c 183 g 231 t
 ORIGIN
 Query Match 28.8%; Score 743; DB 13; Length 827;
 Best Local Similarity 97.68; Pred. No. 4.1e-199;
 Matches 807; Conservative 0; Mismatches 15; Indels 5; Gaps 5;
 QY 45 GAAAGGTCCTCGGGCTGGAGGACCCGCTCCCGGTCGGGAGTCGCTACCTGGGAAGAAG 104
 DB 1 GAAAGGTCCTCGGGCTGGAGGACCCGCTCCCGGTCGGGAGTCGCTACCTGGGAAGAAG 60
 QY 105 AGAATTTCAAGATTAATAAATCTGAACCAACCTTAATGTGCATGATGTCGTCTTAA 164
 DB 61 AGAATTTCAAGATTAATAAATCTGAACCAACCTTAATGTGCATGATGTCGTCTTAA 120
 QY 165 TACAATCTGTTGGAATGACACTGGAGATATATTTATCTGTGCTCAGATGACACCAAAAT 224
 DB 121 TACAATCTGTTGGAATGACACTGGAGATATATTTATCTGTGCTCAGATGACACCAAAAT 180
 QY 225 AGTA-ATTAGTAATCCTTACAGCAGAAAGTTTGAACAACTTGCATGTCGTCAGGGCACCAG 283
 DB 181 AGTAGAGGAGTAATCCTTACAGCAGAAAGTTTGAACAACTTGCATGTCGTCAGGGCACCAG 240
 QY 284 CAAACATATTTAGTCCAAAGTCTTACCTTGACAAATGATAACAGATGTCATCTGCT 343
 DB 241 CAAACATATTTAGTCCAAAGTCTTACCTTGACAAATGATAACAGATGTCATCTGCT 300
 QY 344 CTGGAGATGGAGTAATATTTATACCAAGTTGACCAAGATGACGAAACCAACAGACAAT 403
 DB 301 CTGGAGATGGAGTAATATTTATACCAAGTTGACCAAGATGACGAAACCAACAGACAAT 360
 QY 404 GCCAATTTACGTGTCATTTATGGAACACTTATGAGATGATGACTGACCAATGACCCCTT 463
 DB 361 GCCAATTTACGTGTCATTTATGGAACACTTATGAGATGATGACTGACCAATGACCCCTT 420
 QY 464 ACACATTTCTCTCTGTTGTTGAGTGAAGTGAACCTGTTAGTGGTTGATACACGCATCAAAA 523
 DB 421 ACACATTTCTCTCTGTTGTTGAGTGAAGTGAACCTGTTAGTGGTTGATACACGCATCAAAA 480
 QY 524 CTAGCTGCACAAAAGAAATGTTAAAGATATATTTAATTAAGTCTGACGTGCTGCCA 583
 DB 481 CTAGCTGCACAAAAGAAATGTTAAAGATATATTTAATTAAGTCTGACGTGCTGCCA 540
 QY 584 CCTCTGTTGCTATTTGCCCAACCAATACCATATTTACCTGCTGTTGTTGTTCTGACAGCT 643
 DB 541 CCTCTGTTGCTATTTGCCCAACCAATACCATATTTACCTGCTGTTGTTGTTCTGACAGCT 600
 QY 644 CAGTAGGAA-TATATGATCGGGGAATGCTGGGCACAAAGAGCTACAGGGGAATTTATGAGCT 702
 DB 601 CAGTAGGAACTATATGATCGGGGAATGCTGGGCACAAAGAGCTACAGGGGAATTTATGAGCT 660
 QY 703 CGAGGAGCTACTGGAATGTTGCC-CTTTTATCTCTCCCATCTTAAT-ATAAGTCTCT 760
 DB 661 CGAGGAGCTACTGGAATGTTGCCCGGTTTAAATTCCTTCCCATCTTAATATAAGTCTCT 720

QY 761 GCAGAGTGACATCTCTCTGTGTACAGTGAAGATGTCAG-AGATTCCTCGTTAGTTACTCT 819
 DB 721 GCAGAGTGACATCTCTCTGTGTACAGTGAAGATGTCAG-AGATTCCTCGTTAGTTACTCT 780
 QY 820 TCAGATTACATATATCTTTTGGACCCGAAAGATGATACAGCACGAGA 866
 DB 781 TCAGATTACATATATCTTTTGAACCCGAAAGATGATACAGCACGAGA 827
 RESULT 10
 BF980168/c 951 bp mRNA linear EST 23-JAN-2001
 LOCUS 602288183T1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4373781 3',
 DEFINITION mRNA sequence.
 ACCESSION BF980168
 VERSION BF980168.1 GI:12347383
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 951)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cspas-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10036 row: f column: 22
 High quality sequence start: 22
 High quality sequence stop: 792.
 Location/Qualifiers
 1. .951
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4373781"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTT-3',
 size-selected for average insert size 2.2 kb and
 normalized to 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 204 a 234 c 225 g 288 t
 ORIGIN
 Query Match 28.5%; Score 735.8; DB 12; Length 951;
 Best Local Similarity 95.8%; Pred. No. 4.8e-197;
 Matches 821; Conservative 0; Mismatches 27; Indels 9; Gaps 6;
 QY 1696 ATAACACATCATAGATAAGTTCACAGCCATTTGGATTCCCAATTCAGAGAGAAAT 1755
 DB 850 AAACACACCCGAGAGGTTTCACAGCCAGCCC--TGGATCCACCTCAGAG-AGAAGT 794
 QY 1756 GACCTCAATCTTGATGCTCTTGTGGGGTTCAGAGAAATCTGCTTCATCTGAAAAGCC 1815
 DB 793 GACCTCAATCT--GATCGCTCTGTGGGTCCTCAGAGAAATCTGCTTCATCTGAAAAGCC 737
 QY 1816 AAGGACCAAGAAACT-TCAGATCAGACTAGCAGTGTGTACCAATGAAATACAC 1874
 DB 736 CAGGAACCAAGAAACTCTCAGATCAGACTAGCAGGAGTGCTACCAATGAAAT-ACAC 678


```

QY 2233 ATC 2235
Db 792 TTC 794

RESULT 12
B0433486
LOCUS
DEFINITION
AGENCOURT_7805625 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6067882
5', mRNA sequence.
ACCESSION
B0433486
VERSION
B0433486.1 GI:21172562
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 906)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13348 row: b column: 11
High quality sequence stop: 451.
Location/Qualifiers
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:6067882"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 291 a 216 c 209 g 190 t
ORIGIN
Query Match 28.0%; Score 723; DB 14; Length 906;
Best Local Similarity 96.3%; Pred. No. 2e-193;
Matches 773; Conservative 0; Mismatches 25; Indels 5; Gaps 3;

QY 970 AGCCGGAGAGTGAACGACGAGATGGAGAGCAGAGTCCCAATGTCTCATTTGATGCG 1029
Db 1 AGCCGGAGAGTGAACGACGAGATGGAGAGCAGAGTCCCAATGTCTCATTTGATGCG 60

QY 1030 AGAATGCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTGGAGTTGCACAAAGCAAT 1089
Db 61 AGAATGCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTGGAGTTGCACAAAGCAAT 120

QY 1090 AGAGGACGAGGAAGATCTGACCCAGGAGTGGACAAAGTCAATCAGATATTTCAACTCTT 1149
Db 121 AGAGGACGAGGAAGATCTGACCCAGGAGTGGACAAAGTCAATCAGATATTTCAACTCTT 180

QY 1150 CCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1209
Db 181 CCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 240

QY 1210 CCAGCTGAACAAATTTCTTCAGCCCTTCACATCTTCAATGTCAGCTCAGGCTCATTCG 1269
Db 241 CCAGCTGAACAAATTTCTTCAGCCCTTCACATCTTCAATGTCAGCTCAGGCTCATTCG 300

us-09-781-693a-1.rst
QY 1270 ACATCATCTCCACACAGAAAGCCCTCATTTCTACTCCTTTTGGTATCTTCTCCAGACAGTGAA 1329
Db 301 ACATCATCTCCACACAGAAAGCCCTCATTTCTACTCCTTTTGGTATCTTCTCCAGACAGTGAA 360

QY 1330 CAAAGGCAGTCTGTTGAGGCATCTGGACACACACACATCATCATCATCATCATCATCATCAT 1389
Db 361 CAAAGGCAGTCTGTTGAGGCATCTGGACACACACACATCATCATCATCATCATCATCATCAT 420

QY 1390 GAAAAGCTGAGCCCAAAA--CCAGGACACAGGTGAACACAGTCTTTAAAGTTTTCACACTACAGCA 1447
Db 421 GAAAAGCTGAGCCCAAAA--CCAGGACACAGGTGAACACAGTCTTTAAAGTTTTCACACTACAGCA 480

QY 1448 CAGAAGAACAACTCAAGCACAAATAAACTGAACCTTTTACAGATGAATGAGGACAGATATAG 1507
Db 481 CAGAAGAACAACTCAAGCACAAATAAACTGAACCTTTTACAGATGAATGAGGACAGATATAG 540

QY 1508 CATCAAGTTCTCAGAGAAATTTGGGAGCCATTGCAAAATCTGAGGCTCAGAGGAAATCTTTTCG 1567
Db 541 CATCAAGTTCTCAGAGAAATTTGGGAGCCATTGCAAAATCTGAGGCTCAGAGGAAATCTTTTCG 600

QY 1568 TCCACAGAGCTCAGTGCACACACACAGGAGGAGAGTATCTGCAGAAACCCAGTTTGAGAAC 1627
Db 601 TCCACAGAGCTCAGTGCACACACACAGGAGGAGAGTATCTGCAGAAACCCAGTTTGAGAAC 660

QY 1628 CATCAGAGGATGTGACAAATATATCAGGAAGGAGTATCTGCAGAAACCCAGTTTGAGAAC 1687
Db 661 CATCAGAGGATGTGACAAATATATCAGGAAGGAGTATCTGCAGAAACCCAGTTTGAGAAC 720

QY 1688 ATATCAATAAACAATCAGATAAGTT--CACAGCCCAAGCATTGGATTCCAAAC--TCAG 1744
Db 721 CTTTTCAGTATACCCCATCGGATAAGTTCCCGAGCAAGCCATTGGATTCCCACTCCG 780

QY 1745 GAGAAAGAAATGACCTCAATCTT 1767
Db 781 AAAAAAGAAATGACCTCATCT 803

RESULT 13
AL560929
LOCUS
DEFINITION
AL560929 LTI_NFL010_BC2 722 bp mRNA linear EST 16-FEB-2001
prime, mRNA sequence.
ACCESSION
AL560929
VERSION
AL560929.1 GI:12907861
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 722)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL005Y113"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/notes="vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
FEATURES
source

```


Db 24 AGAGGTTTTGACAAATTCGTTGAGGCGACCGGCAAAACATATTTAGTGCAGAAATCTTT 83
 QY 309 ACCTTGTACAAATGATAAACAGATTGTATCTGCTGCTGGAGTGGAGTAAATATTTATAC 368
 Db 84 ACCTTGTACAAATGATAAACAGATTGTATCTGCTGCTGGAGTGGAGTAAATATTTATAC 143
 QY 369 CAACGTTGAGCAAGATGAGAAACCAACAGACAAATGCAATTTACGCTGCTATTTGGAAC 428
 Db 144 CAACGTTGAGCAAGATGAGAAACCAACAGACAAATGCAATTTACGCTGCTATTTGGAAC 203
 QY 429 TACTTATGAGATTATGACTGTACCAATGACCCCTTACACTTTTCTCTCTGCTGCTGGAAGA 488
 Db 204 TACTTATGAGATTATGACTGTACCAATGACCCCTTACACTTTTCTCTCTGCTGCTGGAAGA 263
 QY 489 TGGAACTGTTAGTGGTTGATACAGCATCAAACTAGCTGCACAAAGAAAGATTGTAA 548
 Db 264 TGGAACTGTTAGTGGTTGATACAGCATCAAACTAGCTGCACAAAGAAAGATTGTAA 323
 QY 549 AGATGATATTTTAACTGTCGACGCTGTCGCCACGCTCTGTTGCTATTTGCCACCAAT 608
 Db 324 AGATGATATTTTAACTGTCGACGCTGTCGCCACGCTCTGTTGCTATTTGCCACCAAT 383
 QY 609 ACCATATTTACCTTCTGCTGTTGCTTCTGACAGCTCAGTACGATATATGATCGGCGAAT 668
 Db 384 ACCATATTTACCTTCTGCTGTTGCTTCTGACAGCTCAGTACGATATATGATCGGCGAAT 443
 QY 669 GCTGGGCAAGAGCTACAGGGAATTTATGAGGTCGAGGCTACTGGAATGGTTGCCCG 728
 Db 444 GCTGGGCAAGAGCTACAGGGAATTTATGAGGTCGAGGCTACTGGAATGGTTGCCCG 503
 QY 729 TTTTATTCCTTCCCATCTTAATAAATGCTGTCGAGAGTGACATCTCTGTGTACAGTGA 788
 Db 504 TTTTATTCCTTCCCATCTTAATAAATGCTGTCGAGAGTGACATCTCTGTGTACAGTGA 563
 QY 789 AGATGCTCAAGAGATTCGTTAGTACTTCTCAGATACATATATCTTTTGACCCGAA 848
 Db 564 AGATGCTCAAGAGATTCGTTAGTACTTCTCAGATACATATATCTTTTGACCCGAA 623
 QY 849 AGATGATACAGCAGAGAACTTTAAACTCTCTTCCGGAAGAGAGAGAGAGTTGGG 908
 Db 624 AGATGATACAGCAGAGAACTTTAAACTCTCTTCCGGAAGAGAGAGAGAGAGTTAGC 683
 QY 909 ACAACCCAGTTAGCGTTGACAGCTTGGTGTGATGTCAGATA 955
 Db 684 GAAAAGCAAGTAAGCGGTGAGAGCTTCTGCTGTGATGTCAGATA 730

RESULT 18
 BG682657 729 bp mRNA linear EST 01-MAY-2001
 LOCUS 602626942F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751578 5',
 DEFINITION mRNA sequence.

ACCESSION BG682657
 VERSION BG682657.1 GI:13914054
 KEYWORDS EST
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 729)

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10608 row: d column: 11

FEATURES
 source

High quality sequence start: 3
 High quality sequence stop: 693.
 Location/Qualifiers
 1..729
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4751578"
 /clone_lib="NCI_CGAP_Skn4"
 /tissue_type="squamous cell carcinoma"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 253 a 165 c 168 g 143 t
 ORIGIN

Query Match 26.3%; Score 678.8; DB 12; Length 729;

Best Local Similarity 99.3%; Pred. No. 6.2e-181;
 Matches 713; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1394 AGCTGAGCCCAACCCAGGACAGGTGAACCAAGTTTAAAGTTTGCACTACAGCAGAG 1453
 Db 12 ATCTGAGCCCAACCA -GGACAGGTGAACCAAGTTTAAAGTTTGCACTACAGCAGAG 70

QY 1454 GAACAACCTACAGCACATTAACCTGAACCTTACAGATGAATGGAGCAGTATAGCATCAA 1513
 Db 71 GAACAACCTACAGCACATTAACCTGAACCTTACAGATGAATGGAGCAGTATAGCATCAA 130

QY 1514 GTTCTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGAATTTTCGTCGCCAC 1573
 Db 131 GTTCTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGAATTTTCGTCGCCAC 190

QY 1574 AGAGCTCAGTCAACCCAGGAGAGACA -GTGAACAAAAGCTCCTGGAAGATCATCA 1632
 Db 191 AGAGCTCAGTCAACCCAGGAGAGACA -GTGAACAAAAGCTCCTGGAAGATCATCA 250

QY 1633 GAGGATGTGACAAAATATCAGGAAGAGTATCTGCAGAAAAGCTTGGAGAACCATATC 1692
 Db 251 GAGGATGTGACAAAATATCAGGAAGAGTATCTGCAGAAAAGCTTGGAGAACCATATC 310

QY 1693 AATATAACAATCAGATAAGTTTCAGCCAAAGCCATTTGGATTCGAATCAGGAGAAAGA 1752
 Db 311 AATATAACAATCAGATAAGTTTCAGCCAAAGCCATTTGGATTCGAATCAGGAGAAAGA 370

QY 1753 AATGACCTCAATCTTGCATCGCTCTTGGGGTTCCAGAGAAGATCTCTTCATCTGAAAAA 1812
 Db 371 AATGACCTCAATCTTGCATCGCTCTTGGGGTTCCAGAGAAGATCTCTTCATCTGAAAAA 430

QY 1813 GCCAAGGAACCAAGAACTTCAGATCAGACTAGCAGTGCAGAGTGCCTACCAATGAAAAATAC 1872
 Db 431 GCCAAGGAACCAAGAACTTCAGATCAGACTAGCAGTGCAGAGTGCCTACCAATGAAAAATAC 490

QY 1873 ACCAATCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAAACA 1932
 Db 491 ACCAATCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAAACA 550

QY 1933 TCCACCAAGGACTCTCTCTTCCAGGACACAGATGACAGTGCATGATGACCCAGTCTGATC 1992
 Db 551 TCCACCAAGGACTCTCTCTTCCAGGACACAGATGACAGTGCATGATGACCCAGTCTGATC 610

QY 1993 CCAGGTGCAAGGTATCAGCAGGACCTGGTGTATGAGAGCTCTGCTGTGGCCGCTATTCAG 2052
 Db 611 CCAGGTGCAAGGTATCAGCAGGACCTGGTGTATGAGAGCTCTGCTGTGGCCGCTATTCAG 670

QY 2053 GAGTTCTTACAGCGGAGAAAGAAAGAAAGAAATGCG -AAGAATTTGGATCTTTGAAAC 2109
 Db 671 GAGTTCTTACAGCGGAGAAAGAAAGAAAGAAATGCG -AAGAATTTGGATCTTTGAAAC 728

RESULT 19
 BI464045
 LOCUS

BI464045 783 bp mRNA linear EST 21-AUG-2001

ORIGIN

Query Match 25.9%; Score 667.8; DB 12; Length 754;
 Best Local Similarity 95.5%; Pred. No. 8.3e-178;
 Matches 698; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 210 AGATGACACCAATAGTAATAGTATCCTTACAGCAGAGAGGTTTGCACAAATTCG 269
 DB 23 AGAGACTGAGAATACCTTTTAAAGGAGAGGTTCTCATCAGAGTTTGCACAAATTCG 82
 QY 270 TTCAGGCGCCGAGCAGCAACATATTAGTGCAGAGTTCTTACCTTGTGCACAAATGATAACA 329
 DB 83 TTCAGGCGCCGAGCAGCAACATATTAGTGCAGAGTTCTTACCTTGTGCACAAATGATAACA 142
 QY 330 GATGTATCTGCTCTGAGATGAGTAATATTATACCAACGTTGAGCAAGATGAGA 389
 DB 143 GACTGTATCTGCTCTGAGATGAGTAATATTATACCAACGTTGAGCAAGATGAGA 202
 QY 390 AACCAACAGACAATGCCAATTTAGCTGTCAATTTATGGAACACTTATGAGATTTAGCTGT 449
 DB 203 AACCAACAGACAATGCCAATTTAGCTGTCAATTTATGGAACACTTATGAGATTTAGCTGT 262
 QY 450 ACCCAATGACCTTACACTTTTCTCTTGTGTTGAAGATGGAACACTTATGAGTGGTTGA 509
 DB 263 ACCCAATGACCTTACACTTTTCTCTTGTGTTGAAGATGGAACACTTATGAGTGGTTGA 322
 QY 510 TACAGCATCAAACTAGCTGCACAAAGAGATTGTAAGATGATATTTTAAATTAACG 569
 DB 323 TACAGCATCAAACTAGCTGCACAAAGAGATTGTAAGATGATATTTTAAATTAACG 382
 QY 570 TCGAGGTGCTGCCAGCTGTGTTGCTATTTGCCACCAATACCATTTACCTTGTGTTGG 629
 DB 383 TCGAGGTGCTGCCAGCTGTGTTGCTATTTGCCACCAATACCATTTACCTTGTGTTGG 442
 QY 530 TTGTTGACAGCTCAGTACAAATATATATGTCGGGATGTCGGCACAAGACTACAGG 689
 DB 443 TTGTTGACAGCTCAGTACAAATATATATGTCGGGATGTCGGCACAAGACTACAGG 502
 QY 690 GAATATGACAGTGCAGGAGTACTGGAATGTTGCCCGTTTTTATCTTCCCTCTTAA 749
 DB 503 GAATATGACAGTGCAGGAGTACTGGAATGTTGCCCGTTTTTATCTTCCCTCTTAA 562
 QY 750 TAAATAGCTGCAGAGTACATCTCTGTGTACAGTGAAGATGTCAGAGATTTCTGT 809
 DB 563 TAAATAGCTGCAGAGTACATCTCTGTGTACAGTGAAGATGTCAGAGATTTCTGT 622
 QY 810 TAGTTACTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGAACT 869
 DB 623 TAGTTACTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGAACT 682
 QY 870 TAAACTCTCTTCTGCGGAGAGAGAGAGAGTTGCGACAAACCACCGATTAAAGCGTTT 929
 DB 683 TAAACTCTCTTCTGCGGAGAGAGAGAGAGTTGCGACAAACCACCGATTAAAGCGTTT 741
 QY 930 GAGACTTCGTG 940
 DB 742 GAGACTTCGTG 752

RESULT 21
 BI461627
 LOCUS 834 bp mRNA linear EST 21-AUG-2001
 DEFINITION 603206744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5272379 5',
 mRNA sequence.
 ACCESSION BI461627
 VERSION BI461627.1 GI:15252283
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 834)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLML1687 row: h column: 12
 High quality sequence start: 5
 High quality sequence stop: 725.

FEATURES
source

1..834
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5272379"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 244 a 163 c 193 g 234 t

Query Match 25.8%; Score 665.8; DB 13; Length 834;
 Best Local Similarity 95.2%; Pred. No. 3.2e-177;
 Matches 719; Conservative 0; Mismatches 32; Indels 4; Gaps 3;

QY 249 AAAGTTTTGACACAATTCGTTTCAGGGCAGCGAGCAACATATTTAGTGCAGAGTTCTT 308
 DB 24 AGAGTTTTGACACAATTCGTTTCAGGGCAGCGAGCAACATATTTAGTGCAGAGTTCTT 83
 QY 309 ACCTTGTAACAATGATAACAGATTGTATCTGCTCGAGATGGAGTAATATTTATATAC 368
 DB 84 ACCTTGTAACAATGATAACAGATTGTATCTGCTCGAGATGGAGTAATATTTATATAC 143
 QY 369 CAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTAGTGTCAATATGGAAC 428
 DB 144 CAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTAGTGTCAATATGGAAC 203
 QY 429 TACTTATGAGATTATGACTGTACCAATGACCTTACACTTTCTCTCTGTTGGTGAAGA 488
 DB 204 TACTTATGAGATTATGACTGTACCAATGACCTTACACTTTCTCTCTGTTGGTGAAGA 263
 QY 489 TGGAACTGTTAGTGGTTGTATACAGCATCAAACTAGCTGCACAAAAGAGATTGTA 548
 DB 264 TGGAACTGTTAGTGGTTGTATACAGCATCAAACTAGCTGCACAAAAGAGATTGTA 323
 QY 549 AGATGATATTTTAAATTAAGTGTGCTGCCAGTCTGTTGCTATTTGTCGCCCAAT 608
 DB 324 AGATGATATTTTAAATTAAGTGTGCTGCCAGTCTGTTGCTATTTGTCGCCCAAT 383
 QY 609 ACCATATACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 668
 DB 384 ACCATATACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 443
 QY 669 GCTGGGCAAGAGCTACAGGAATTTATGAGTGCAGGAGTACTGGAATGTTGCCCG 728
 DB 444 GCTGGGCAAGAGCTACAGGAATTTATGAGTGCAGGAGTACTGGAATGTTGCCCG 503
 QY 729 TTTTATCTTCCCATCTTAAATTAAGTGTGCTGCCAGAGTGACATCTCTGTTACAGTGA 788
 DB 504 TTTTATCTTCCCATCTTAAATTAAGTGTGCTGCCAGAGTGACATCTCTGTTACAGTGA 563


```

QY 789 AGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTTGGACCCGAA 848
DB 564 AGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTTGGACCCGAA 623
QY 849 AGATGATACAGACAGAGACTTAAACTCTCTCTCGGGAAGAG-AGAAGAGAGAGTTGC 907
DB 624 AGATGATCCGACGAGAACTTAAACTCTCTCTCGGGAAGAGCAGAAAGAGAGTTGC 683
QY 908 G-ACAACCACCAGTTTAAAGGTTTGAGACTTCGTGTGA--TTGGTCAGATACCTGGACCCA 964
DB 684 GAACAACCACCAGTTTAAAGGTTTGAGACTTCGTGTGA--TTGGTCAGATACCTGGACCCA 743
QY 965 GAGCAGGCCGAGAGTGAACGAGACGAGATGGA 999
DB 744 GAAGCATGCCGGAAGTACCAGAGACGAGAAGGA 778

RESULT 22
LOCUS B1464622 850 bp mRNA linear EST 21-AUG-2001
DEFINITION 603203236f1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268920 5',
ACCESSION B1464622
VERSION B1464622.1 GI:15255278
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LM11678 row: h column: 09
High quality sequence stop: 710.
FEATURES
source
location/Qualifiers
1..850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5268920"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTIVN-3',
size-selected for average insert size 2.2 kb and
normalized to R0.5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 243 a 182 c 188 g 237 t
ORIGIN
Query Match 25.7%; Score 663.6; DB 13; Length 850;
Best Local Similarity 93.0%; Pred. No. 1.4e-176;
Matches 785; Conservative 0; Mismatches 44; Indels 15; Gaps 8;
QY 1 ATGCTCGGGGTGGCTCTACCCACACCTGTGTGGACGTGAGAAAGTCCCTCGGG 60
DB 6 ATGCTCGGGGTGGCTCTACCCACACCTGTGTGGACGTGAGAAAGTCCCTCGGG 65

```

```

61 CTGAGGACCGCTCCCGCTGGGAGTCGCTACCTGGGGAAGAGAAATTTATCCAAAGA 120
66 CTGAGGACCGCTCCCGCTGGGAGTCGCTACCTGGGGAAGAGAAATTTATCCAAAGA 125
121 TTAATACTTGAGCAACCCCTTAATGTGCATGATGGTGTGTTAATAACATCTGTTGGAAT 180
126 TTAATACTTGAGCAACCCCTTAATGTGCATGATGGTGTGTTAATAACATCTGTTGGAAT 185
181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 240
186 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 245
241 TACAGCAAAAGGTTTTCACAACAATTCGTTCCAGGSCCGGAGCAACATATTAGTGCA 300
246 TACAGCAAAAGGTTTTCACAACAATTCGTTCCAGGSCCGGAGCAACATATTAGTGCA 305
301 AAGTTCCTTACCTTGTCACAAATGATAAAGATGTTATCTCTCTGGAGATGAGTAATA 360
306 AAGTTCCTTACCTTGTCACAAATGATAAAGATGTTATCTCTCTGGAGATGAGTAATA 365
361 TTTTATACCAACGTTGAGCAAGATCCAGAAACCAACACAGCAATGCCAATTTAGTGCA 420
366 TTTTATACCAACGTTGAGCAAGATCCAGAAACCAACAGCAATGCCAATTTAGTGCA 425
421 TATGGAACCTACTTATGAGATTATGACTGTACCAATGACCCCTTACACTTTCTCTCTCT 480
426 TATGGAACCTACTTATGAGATTATGACTGTACCAATGACCCCTTACACTTTCTCTCTCT 485
481 GGTGAAGATGGAATCTGTAGTGGTTGATACACGCATCAAAACTAGTGCACAA--A 536
486 GGTGAAGATGGAATCTGTAGTGG--TTGATACACGCATCAAAACTAGTGCACAAACGA 544
537 AGAAGATGTTAAAGATGATATTTAATAA--TGTCCAGCTGCTGCCACGCTCTGCT 594
545 AGATGTTACACAGATGATCTTTAATAA--TGTCCAGCTGCTGCCACGCTCTGCT 604
595 ATTGCCCCCAATACCATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 654
605 ATTGCCCCCAATACCATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 664
655 TATGATCGGGGAATGCTGGGCAC--AAGAGCTACAGGG--AATTATGCAGG--TCGAGGGAC 710
665 TATGATCGGGGAATGCTGGGCAC--AAGAGCTACAGGG--AATTATGCAGG--TCGAGGGAC 724
711 -TACTGGAATGGTGGCCGCTTTTATTCCTTCCCATCTTAATAAAGCTCCGACAGTCA 769
725 ATACTGGAATGGTGGCCGCTTTTATTCCTTCCCATCTTAATAAAGCTCCGACAGTCA 784
770 CATCTCTGTTTACA---GTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATT 826
785 ACTCTCTGTTTACACACTGCAGATGGTCAACAAATCCTCGTAATTCACCTCCCATTTAT 844
827 ACAT 830
845 ACTT 848

RESULT 23
LOCUS BE180558 694 bp mRNA linear EST 22-JUN-2000
DEFINITION RC3-H70625-130400-021-c02 H70625 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE180558
VERSION BE180558.1 GI:8659734
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 694)
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

```


M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=st2-RC3-HT0625-130400-021-c02&t3=2000-04-13&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 678.

FEATURES		Location/Qualifiers	
Source	1.. 694	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone_lib="Hr0625"	
		/dev_stage="Adult"	
		/note="Organ: head_neck; Vector: puc18; Site_1: Smal; Site_2: SmaI; A mini-library was made by cloning products derived from ORESPEC PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	208 a	136 c	207 t
ORIGIN			
	Query Match	25.5%;	Score 658.2; DB 10; Length 694;
	Best Local Similarity	99.3%;	Pred. No. 4.2e-175;
	Matches 682; Conservative	0;	Mismatches 3; Indels 2; Gaps 2;
QY	187	GGAGAATATATTTTATCTGGCTCAGATGACACAAATTAGTAATTAGTAAATCCTTACAGC	246
Db	1	GGAGATATATTTTATCTGGCTCAGATGACACCAATTAGTAATTAGTAAATCCTTACAGC	60
QY	247	AGAAAGGTTTGGACAAATTCGTTGAGGCGACGAGCAACATATATTTAGTGCAAGTTC	306
Db	61	AGAAAGG-TTTGACAAATTCGTTGAGGCGACGAGCAACATATCTAGTGCAAGTTC	119
QY	307	TTACCTTGTCAAATGATAACAGATTGTATCCTGCTCTGGAGATGAGTAATATTTTAT	366
Db	120	TTACCTTGTCAAATGATAACAGATCGTATCTGCTCTGGAGATGAGTAATATTTAT	179
QY	367	ACCAACGTTGAGCAAGATGAGAAACCAACAGACAATGCCAATTTAGCTGTCATATTGGA	426
Db	180	ACCAACGTTGAGCAAGATGAGAAACCAACAGACAATGCCAATTTAGCTGTCATATTGGA	239
QY	427	ACTACTTATGAGATPATGACTGTATCCCAATGACCCCTTTACACTTTTCTCTGTGGTGAA	486
Db	240	ACTACTTATGAGATPATGACTGTATCCCAATGACCCCTTTACACTTTTCTCTGTGGTGAA	299
QY	487	GATGGAACCTGTTAGTGGTTTGATACGCGATCAAACTAGCTGCACAAAAGAAGATTGT	546
Db	300	GATGGAACCTGTTAGTGGTTTGATACGCGATCAAACTAGCTGCACAAAAGAAGATTGT	359
QY	547	AAAGATGATATTTTAAATTAACGTGCGAGCTGCTGCCACGTCTGTGTGCTATTTGCCACCA	606
Db	360	AAAGATGATATTTTAAATTAACGTGCGAGCTGCTGCCACGTCTGTGTGCTATTTGCCACCA	419
QY	607	ATACCATATTACCTTGCTGTGGTTGTTCTGACAGCTCAGTACGAATATATGATCGGGGA	666
Db	420	ATACCATATTACCTTGCTGTGGTTGTTCTGACAGCTCAGTACGAATATATGATCGGGGA	479
QY	667	ATGTGGGCCACAAGACTACAGGGAAATTATCGAGTTCGAGGACTACTGGAAATGTTGCC	726

Db	480	ATGCTGGGCACAGAGCTACAGGGAATTATGACGGTCTGAGGACTACTGGAATGGTTGCC	539
Qy	727	CGTTTATTCCTCCCATCTTAATAAGTCCGTCAGAGTGACATCTCTGTGTACAGT	786
Db	540	CGTTTATTCCTCCCATCTTAATAAGTCCGTCAGAGTGACATCTCTGTGTACAGT	599
Qy	787	GAGATGCTCAAGAGATCTCGTTAGTTACTTTCAGATTACATATATCTTTTGGACCG	846
Db	600	GAGATGCTCAAGAGATCTCGTTAGTTACTTTCAGATTACATATATCTTTTGGACCG	659
Qy	847	AAAGATGATACAGCAGGAGACTTAAA	873
Db	660	AAAGATGATACAGCAGGAGACTTAAA	685
RESULT 24			
LOCUS	BG117357	602346361F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4441431	5',
DEFINITION	mRNA sequence.		
ACCESSION	BG117357	GI:12610863	
VERSION	BG117357.1	EST	
KEYWORDS	human.		
SOURCE	ORGANISM Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 739)		
	NIH-MGC http://mgi.nci.nih.gov/.		
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-r@mail.nih.gov		
JOURNAL	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
COMMENT	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM10212 row: 1 column: 16		
FEATURES	High quality sequence stop: 657.		
	Location/Qualifiers		
	1..739	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
source		/clone="IMAGE:4441431"	
		/clone_lib="NIH_MGC_90"	
		/tissue_type="adenocarcinoma, cell line"	
		/lab_host="DH10B (phage-resistant)"	
BASE COUNT	222 a	143 c	177 g
	Note: this is a NIH_MGC Library.		
	Site_2: Sali; Cloned unidirectionally; oligo-dT primed.		
	Average insert size 1.7 kb. Library enriched for		
ORIGIN	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH_MGC Library.		
	Query Match 25.4%; Score 655.8; DB 12; Length 739;		
	Best Local Similarity 97.7%; Pred. No. 2.1e-174;		
Qy	538	GAAGATTGTAAGATCATATTTTAACTCTGACGCTGCCAGCTCTGTGTCTATT	597
	1	GAAGATTGTAAGATCATATTTTAACTCTGACGCTGCCAGCTCTGTGTCTATT	60
	598	TGCCACCAATACCATATTTACCTTGTGTGTGTGTCTGACAGCTCAGTACGAATATAT	657
	61	TGCCACCAATACCATATTTACCTTGTGTGTGTGTCTGACAGCTCAGTACGAATATAT	120
Db	658	GATCGCGCAATGCTGGGCAAGAGCTACAGGAATATTCAGGTCGAGGACTACTGGA	717

```

Db 121 GATCGCGGAATGCTGGGACACAGAGCTACAGGAATTAATGAGGTGCGAGGACTACTGGA 180
QY 718 ATGGTTGCCGTTTATTCCTTCCATCTTAATAAAGTCTCTGAGAGTGAGCATCTCTG 777
Db 181 ATGGTTGCCGTTTATTCCTTCCATCTTAATAAAGTCTCTGAGAGTGAGCATCTCTG 240
QY 778 TCTTACAGTGAAGTGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATTACATATCTT 837
Db 241 TGTTCAGTGAAGTGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATTACATATCTT 300
QY 838 TTTGACCGAAAGATGATACACAGAGAGTCTTAAACCTCTCTGCGGAAGAGAGA 897
Db 301 TTTGACCGAAAGATGATACACAGAGAGTCTTAAACCTCTCTGCGGAAGAGAGA 360
QY 898 GAAGAGTTCGACACACAGAGTCTTAAAGCTTTGAGACTTCGTTGGTGTGATGTCAGATACT 957
Db 361 GAAGAGTTCGACACACAGAGTCTTAAAGCTTTGAGACTTCGTTGGTGTGATGTCAGATACT 420
QY 958 GGACCCAGAGCAAGCGGAGAGTCAAGCAGAGAGATGGAGACAGATGCCAATGTG 1017
Db 421 GGACCCAGAGCAAGCGGAGAGTCAAGCAGAGAGATGGAGACAGATGCCAATGTG 480
QY 1018 TCATTGATGACAGAGATCTCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAGTT 1077
Db 481 TCATTGATGACAGAGATCTCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAGTT 540
QY 1078 GCACAAAGCAATAGAGACGAGAGAGATCTCGACCCAGAGTGGAAACCAATCAGAT 1137
Db 541 GCACAAAGCAATAGAGACGAGAGAGATCTCGACCCAGAGTGGAAACCAATCAGAT 600
QY 1138 ATTTCAACTCTTCTACCGTCCCATCAAGTCTGATTTGGAAGTCAAGTCAAGTCAAGT 1197
Db 601 ATTTCAACTCTTCTACCGTCCCATCAAGTCTGATTTGGAAGTCAAGTCAAGTCAAGT 658
QY 1198 GAAGTAGATCTCAGCTGAACAATTTCTTCCAGCTTCTACATCTCTTACATCTCAGCT 1257
Db 659 GACGTAAATACCTCCAG-TGAACAATTTCTTCCAGC--TTCTAATCTCTCAAGAGTCAAGT 715
QY 1258 CAGGC 1262
Db 716 AAGGC 720

RESULT 25
BM923018
LOCUS
DEFINITION
AGENCOURT_6632140 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5756969
5', mRNA sequence.
ACCESSION
BM923018
VERSION
BM923018.1 GI:19373397
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
NTH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@r-email.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone Distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium information can be
http://image.lnl.gov
plate: L1A12798 row: g column: 18
High quality sequence start: 14
High quality sequence stop: 693.
Location/Qualifiers
1. .1075

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5756969"
/clone_lib="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH108"
/notes="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(deprecated); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH_MGC Library."
BASE COUNT 277 a 264 c 267 g 267 t
ORIGIN

```

```

Query Match 25.1%; Score 648.6; DB 14; Length 1075;
Best Local Similarity 95.8%; Pred. No. 2.8e-172;
Matches 731; Conservative 0; Mismatches 24; Indels 8; Gaps 6;

QY 1 ATGCTCTCGGGTGGCTCTTACCCACACCTGTTGGGAGCGTGAGGAAAGTCCCTCGG 60
Db 226 ATGCTCTCGGGTGGCTCTTACCCACACCTGTTGGGAGCGTGAGGAAAGTCCCTCGG 285
QY 61 CTGGAGGACCCGCTCCCGCTCGGAGTCGCTACCTCGGAGAGAGAGAAATTTATCCAAAGA 120
Db 286 CTGGAGGACCCGCTCCCGCTCGGAGTCGCTACCTCGGAGAGAGAGAAATTTATCCAAAGA 345
QY 121 TTAACAACCTTGAGCAACCCCTTAATGTCATGATGTTGTTAAATACAACTCTGTTGGAAT 180
Db 346 TTAACAACCTTGAGCAACCCCTTAATGTCATGATGTTGTTAAATACAACTCTGTTGGAAT 405
QY 181 GACACTGAGAGATATATTTATCTGGCTCAGATGACACCAAAATTTAGTAATCT 240
Db 406 GACACTGAGAGATATATTTATCTGGCTCAGATGACACCAAAATTTAGTAATCT 465
QY 241 TACAGCAAAAGTGTTCGACAACTTCGTTCCAGGCGCAGGCAACCAATTTAGTGA 300
Db 466 TACAGCAAAAGTGTTCGACAACTTCGTTCCAGGCGCAGGCAACCAATTTAGTGA 525
QY 301 AAGTTCTTACCTTGTAACAAATGATAACAGATTTGATCTCTGAGATGGAGTAATA 360
Db 526 AAGTTCTTACCTTGTAACAAATGATAACAGATTTGATCTCTGAGATGGAGTAATA 585
QY 361 TTTATACCAACGTTGACCAAGATCGCAAAACCAACACAGCAATGCAATTTACGTGTAT 420
Db 586 TTTATACCAACGTTGACCAAGATCGCAAAACCAACAGCAATGCAATTTACGTGTAT 645
QY 421 TATGGAACCTACTTATGAGATTATGACTGTACCCAAATGA-CCCTTACACTTTTCTCTCTTG 479
Db 646 TATGGAACCTACTTATGAGATTATGACTGTACCCAAATGACCCCTTACACTTTTCTCTCTTG 705
QY 480 TGGTGAAGATGCAACTGTTAGGTGTTGATACAGCATCAAAACTAGCTGCACAAAAGA 539
Db 706 TGGTGAAGATGCAACTGTTAGGTGTTGATACAGCATCAAAACTAGCTGCACAAAAGA 765
QY 540 AGATTGTAAGATGATATTTAAATTAAGTGTGAGGTCGCTGCCAGCTGTGTTGCTATTG 599
Db 766 AGATTGTAAGATGATATTTAAATTAAGTGTGAGGTCGCTGCCAGCTGTGTTGCTATTG 825
QY 600 CCCACCAATACCATATTTACCTTGTGTTGTTT-GTTCTGACAGCTCAGTACGATATATG 658
Db 826 CCCACCAATACCATATTTACCTTGTGTTGTTTGTCTGAGACGCTCCGTCAGATATATG 885
QY 659 AT-CGGGAATGCTGGGACAAAG-AGCTACAGGGAATTTATCCAGGTGCGA--GGGACTACT 714
Db 886 ATCCGGGAATGCTGGGACAAAGACCTCCCGGGAATTTATCCAGGCGCAAGGGAATACCT 945
QY 715 GGAATGCTG--CCCGTTTATCTCTCCCATCTTATAATAA 755
Db 946 GGAATGCTGCGCGGTTTAAATCCCTTCCCATCTTATAATAA 988

```

```

RESULT 26
BE867206      775 bp      mRNA      linear      EST 20-OCT-2000
LOCUS         601442266F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846472 5',
DEFINITION   mRNA sequence.
ACCESSION    BE867206
VERSION      BE867206.1 GI:10315982
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    NIH-MGC http://mgc.nci.nih.gov/
AUTHORS      1 (bases 1 to 775)
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM9559 row: c column: 17
              High quality sequence stop: 611.
              Location/Qualifiers
                1..775
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3846472"
                  /clone_lib="NIH_MGC_65"
                  /tissue_type="adenocarcinoma"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                  Average insert size 1.8 kb. Library constructed by Life
                  Technologies."
BASE COUNT   255 a 182 c 186 g 152 t
ORIGIN
Query Match      25.1%; Score 646.6; DB 12; Length 775;
Best Local Similarity 94.4%; Pred. NO. 8.6e-172;
Matches 694; Conservative 0; Mismatches 34; Indels 7; Gaps 2;

QY  909 ACAACCCAGTTAAGCGTTTCAGACTTCGTGGTATGGTCAGATCTGGACCCAGAGC 968
      |||||||
DB   1 ACACCCAGTTAAGCGTTTCAGACTTCGTGGTATGGTCAGATCTGGACCCAGAGC 60

QY  969 AAGCGCGGAGAGTGAACGAGACGAGATGGAGCAGAGTCCCAATGTCTATTGATGCA 1028
      |||||||
DB   61 AAGCGCGGAGAGTGAACGAGACGAGATGGAGCAGAGTCCCAATGTCTATTGATGCA 120

QY  1029 GAGATGTCTGATATGTTATCAAGATGGTTTGAAGACGAAGTGGTTGCAACAAGCAA 1088
      |||||||
DB   121 GAGATGTCTGATATGTTATCAAGATGGTTTGAAGACGAAGTGGTTGCAACAAGCAA 180

QY  1089 TAGAGCAGCAGGAAGATCTCGACCCAGAGTGGACAAAGTCAATCAGATATTTCAACTCT 1148
      |||||||
DB   181 TAGAGCAGCAGGAAGATCTCGACCCAGAGTGGACAAAGTCAATCAGATATTTCAACTCT 240

QY  1149 TCCTACGGTCCCATCAAGTCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1208
      |||||||
DB   241 TCCTACGGTCCCATCAAGTCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300

QY  1209 TCCAGTGAACAATTTCTCAGCCTTCTACATCCTCTACAAATGTCAAGTCAAGTCAAGTCAATC 1268
      |||||||
DB   301 TCCAGTGAACAATTTCTCAGCCTTCTACATCCTCTACAAATGTCAAGTCAAGTCAAGTCAATC 360

QY  1269 GACATCATCTCCACAGAAAGCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1328

```

```

|||||
Db  361 GACATCATCTCCACAGAAAGCCCTCATCTTACCTTGTCTATCTCTCCAGACAGTGA 420

QY  1379 ACAAGGCGAGTCTGTGAGGCATCTGGACACACACACATCATCATGCTGTGATAACAATAA 1388
      |||||||
Db  421 ACAAGGCGAGTCTGTGAGGCATCTGGACACACACACATCATCATGCTGTGATAACAATAA 480
      |||||||
QY  1389 TGAAGAAGCTGAGCCGCCAACACAGGAGTGAACAGTCTTTAAAGTTTGCACACTACAGCAC 1448
      |||||||
Db  481 TGAAGAAGCTGAGCCGCCAACACAGGAGTGAACAGTCTTTAAAGTTTGCACACTACAGCAC 540
      |||||||
QY  1449 AGAGGGAACAACTCAACACCAATAAATAAAGTGAACATTTACAGATGAATGGAGCAGTATAGC 1508
      |||||||
Db  541 AGAGGGAACAACTCAACACCAATAAATAAAGTGAACATTTACAGATGAATGGAGCAGTATAGC 600
      |||||||
QY  1509 ATCAAGTTTCTAGAGAAATT--GGGAGCCATTGCAATCTGAGGGTCAGGAGG-----AAT 1561
      |||||||
Db  601 ATCAAGTTTCTAGAGAAATCGGGAAGCCATTGCAATCTGAGGGTCAGGAGGGAATCTAC 660
      |||||||
QY  1562 CTTTGTGTCACAGAGTCTAGTCAACACCCAGAGGAGACAGTGAACAAAGCTCTCTG 1621
      |||||||
Db  661 CGTCCCAACAGAGCTCAGAGGGAACACGACGAGAGGAGCCCGGAACAAAGGCGCTG 720
      |||||||
QY  1622 AAGATCATCAGAGG 1636
      |||||||
Db  721 GAGAACCCACCAAG 735

RESULT 27
LOCUS         BG772319      806 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION   602722419F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4839173 5',
              mRNA sequence.
ACCESSION    BG772319
VERSION      BG772319.1 GI:14082972
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    NIH-MGC http://mgc.nci.nih.gov/
AUTHORS      1 (bases 1 to 806)
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
              cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
              Toshiyuki and Piero Carninci (RIKEN)
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM10774 row: f column: 06
              High quality sequence stop: 791.
              Location/Qualifiers
                1..806
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4839173"
                  /clone_lib="NIH_MGC_97"
                  /lab_host="DH10B"
                  /note="Organ: testis; Vector: pBluescriptR (modified
                  pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag
                  ); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
                  size-selected for average insert size 2.2 kb and
                  normalized to 5. This is a primary library enriched
                  for full-length clones and constructed using the
                  Cap-trapper method (Carninci, in preparation). Library
                  constructed by M. Brownstein (NIH/NHGRI, National
                  Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT   230 a 169 c 181 g 226 t

```

ORIGIN

Query Match 24.9%; Score 641.8; DB 12; Length 806;
 Best Local Similarity 99.1%; Pred. No. 2e-170; Indels 4; Gaps 4;
 Matches 687; Conservative 0; Mismatches 2;

QY 97 GAAGAAGAGAGATTTATCAAGAGATTTAAACCTTGAAGCAACCCCTTAATGTGCATGATGGT 156
 Db 105 GGAAGAAGAGAGATTTATCAAGAGATTTAAACCTTGAAGCAACCCCTTAATGTGCATGATGGT 164

QY 157 TGTGTTAATACATCTGTTGGATGACACTGGAGATATATTTATCTGCTCAGATGAC 216
 Db 165 TGTGTTAATACATCTGTTGGATGACACTGGAGATATATTTATCTGCTCAGATGAC 224

QY 217 ACCAAATAGTATAGTAAATCCTTACAGCAGAAAGGTTTTGACAACAATTCGTTTCAGGG 276
 Db 225 ACCAAATAGTATAGTAAATCCTTACAGCAGAAAGGTTTTGACAACAATTCGTTTCAGGG 284

QY 277 CACCAGCAACATATTTAGTGCACAAAGTCTTTACCTTGACAAATGATAAAGAGATGTA 336
 Db 285 CACCAGCAACATATTTAGTGCACAAAGTCTTTACCTTGACAAATGATAAAGAGATGTA 344

QY 337 TCTGCTCTGGAGATGGATATATTTATACCAACGTTGAGCAAGATGACAGAAACCCAC 396
 Db 345 TCTGCTCTGGAGATGGATATATTTATACCAACGTTGAGCAAGATGACAGAAACCCAC 404

QY 397 AGACAAATGCCAATTTACGTGTCAATATGGAATCTACTATGAGATATGACTGTACCCAA 456
 Db 405 AGACAAATGCCAATTTACGTGTCAATATGGAATCTACTATGAGATATGACTGTACCCAA 464

QY 457 GACCCTTACATCTTCTCTCTGTTGTCGAAGATGCAACTGTAGTGTGTTTATACAGCG 516
 Db 465 GACCCTTACATCTTCTCTCTGTTGTCGAAGATGCAACTGTAGTGTGTTTATACAGCG 524

QY 517 ATCAAACTAGTGCACAAAAGAGATGTAATGATGATATTTAACTATGCTGACG 575
 Db 525 ATCAAACTAGTGCACAAAAGAGATGTAATGATGATATTTAACTATGCTGACG 584

QY 576 TGCTGCCACGCTGTGCTATTTCCCAACCAATACCATATTAACCTGTGTTGTTTC 635
 Db 585 TGCTGCCACGCTGTGCTATTTCCCAACCAATACCATATTAACCTGTGTTGTTTC 644

QY 636 TGACAGCTAGTACGAATATATGATCGCGCAATGCTGGCAAGAGCTACAGGGAATTA 695
 Db 645 TGACAGCTAGTACGAATATATGATCGCGCAATGCTGGCAAGAGCTACAGGGAATTA 704

QY 696 TGCAGGTCGAGGACTACTGGAA-TGGTGGCCCGTTTATTCCTTCCC-ATCTTAAAT 753
 Db 705 TGCAGGTCGAGGACTACTGGAA-TGGTGGCCCGTTTATTCCTTCCC-ATCTTAAAT 764

QY 754 AAGTCCTG-CAGAGTGACATCTCTGTGTTACAG 785
 Db 765 AAGTCCTGCCAGAGTGACATCTCTGTGTTACAG 797

RESULT 28
 BG530857
 LOCUS 60255984F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:469798 5',
 DEFINITION mRNA sequence.
 ACCESSION BG530857
 VERSION BG530857.1 GI:13522394
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 911)
 NIH-MGC <http://mgi.nci.nih.gov/>
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1527 row: k column: 23
 High quality sequence stop: 701.

FEATURES
source

1. .911
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:469798"
 /tissue_lib="NIH_MGC_61"
 /tissue_type="embryonal carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccattggcc); Site_2: SfiI (ggccattggcc);
 Double-stranded cDNA was prepared from cell line RNA.
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATATGCGC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGGCGGCGACATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library."

BASE COUNT 310 a 168 c 214 g 219 t
 ORIGIN

Query Match 24.9%; Score 641.6; DB 12; Length 911;
 Best Local Similarity 90.1%; Pred. No. 2.5e-170;
 Matches 762; Conservative 0; Mismatches 34; Indels 50; Gaps 5;

QY 1676 CAGTTGAGAACCATATCAATATACACAAATCAGATTAAGTTTCACAGCAAGCCATGGATT 1735
 Db 1 CAGTTGAGAACCATATCAATATACACAAATCAGATTAAGTTTCACAGCAAGCCATGGATT 60

QY 1736 CCAACTCAGGAGAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAAT 1795
 Db 61 CCAACTCAGGAGAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAAT 120

QY 1796 CTGCTTCATCTGAAAAAGCCAAAGAACCAAGAACCTTTCAGATCAGACTAGCAGATG 1855
 Db 121 CTGCTTCATCTGAAAAAGCCAAAGAACCAAGAACCTTTCAGATCAGACTAGCAGATG 180

QY 1856 CTACCAATGAAAAATTAACCAATCTGAGCCCTCAGTTCCAAACAGAGCCACTGGGCCTT 1915
 Db 181 CTACCAATGAAAAATTAACCAATCTGAGCCCTCAGTTCCAAACAGAGCCACTGGGCCTT 240

QY 1916 CAGCTCATGAAGAAACATCCACAGGAGACTCTGCTTTCAGGACACAGATGACAGTATG 1975
 Db 241 CAGCTCATGAAGAAACATCCACAGGAGACTCTGCTTTCAGGACACAGATGACAGTATG 300

QY 1976 ATGACCCAGTCTCTGATCCCAAGGTATCGAGCAGGACCTGGTGTATAGACGCTCTG 2035
 Db 301 ATGA-----TGATAGACGCTCTG 318

QY 2036 CTGTTGCCCGTATTCAGGAGTTCTTCAGAGGAGAAAGAAAGAAATGGAAGAAAT 2095
 Db 319 CTGTTGCCCGTATTCAGGAGTTCTTCAGAGGAGAAAGAAAGAAATGGAAGAAAT 378

QY 2096 TGGATACCTTTGAACATTAGAAGCGCGCTAGTAAATAATGGTTTATAAAGCCATCGCAACT 2155
 Db 379 TGGATACCTTTGAACATTAGAAGCGCGCTAGTAAATAATGGTTTATAAAGCCATCGCAACT 438

QY 2156 CAGGACCAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTATAGTGGTCTG 2215
 Db 439 CAGGACCAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTATAGTGGTCTG 498

QY 2216 ACTGTGCCACATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAG 2275

|||||
Db 499 ACTGTGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGATGCTTCTGGAAG 558
QY 2276 CTGATAATCATGTGGTAAATCGCTGAGCCACATCGTTTGACCCCAATTT-TAGCCCTCA 2334
Db 559 CTGATAATCATGTGGTAAATCGCTGAGCCACATCGTTTGACCCCAATTTTGATGCTTCTGGAAG 618
QY 2335 TCTGGCATAGATTATGACATAAAGATCTGTGTCACCA- - -TTAGAAGAGTCAAGGATTTT 2390
Db 619 TCTGGCATAGATTATGACATAAAGATCTGTGTCACCAATATAGAAGTCAAGGATTTT 678
QY 2391 TAACCGAAATCTGCTGATGAAGTT-ATAACTCGAAGCAACTCATGCTGGAAGAACTA 2449
Db 679 TAACCGAAATCTGCTGATGAAGTTAATAAGTCGAAGCAACTCATGATGGAAGAACTA 738
QY 2450 GAACACCAATACAGTTCCAGCCCTTTCATGTTGAGGATGTTGCTTCACTTAATCATTA 2509
Db 739 GAACCACT-ATTAAGTTCCAAAGCATATTATGTTGAGGATGTTGGGTTCACTTAATCATTA 796
QY 2510 TCCGAG 2515
Db 797 TTCGAG 802

RESULT 29
LOCUS BQ019842/c
DEFINITION BQ019842 648 bp mRNA linear EST 27-MAR-2002
IMAGE:5827567 3', mRNA sequence.
ACCESSION BQ019842
VERSION BQ019842.1 GI:19755119
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 648)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbcr@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M13 FORWARD
POLYA=Yes.

Location/Qualifiers
1. 648
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5827567"
/clone_lib="NCI_CGAP_ED0"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pubic Bone; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site 1: EcoR I;
Site 2: Not I; NCI_CGAP_ED0 is a cDNA library containing
the following tissue(s): Chondrosarcoma cell line C85. The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this

Library is GCTCAAGGCT.
TAG_LIB=UI-H-ED0
TAG_TISSUE=chondrosarcoma
TAG_SEQ=CTCAAGGCT*
BASE COUNT 123 a 143 c 148 g 234 t
ORIGIN
Query Match 24.8%; Score 640; DB 14; Length 648;
Best Local Similarity 99.2%; Pred. No. 5.9e-170;
Matches 643; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1438 CACTACAGCAGAGGAGCAACATACAGACACATAAATCTTACAGATGAATGG 1497
Db 648 CACTACAGCAGAGGAGCAACATACAGACACATAAATCTTACAGATGAATGG 589
QY 1498 AGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCCAATTCGAAATCTGAGGGTCAGAG 1557
Db 588 AGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCCAATTCGAAATCTGAGGGTCAGAG 529
QY 1558 GAATCTTTCTGTCACAGAGCTCAGTGAACACACAGAGGAGGAGCAAGTGAACAAAGCT 1617
Db 528 GAATCTTTCTGTCACAGAGCTCAGTGAACACACAGAGGAGGAGCAAGTGAACAAAGCT 469
QY 1618 CCTGAAGAAATCATCAGAGGATGTGACAAATATATCAGGAGGAGTATCTGCAGAAACCCA 1677
Db 468 CCTGAAGAAATCATCAGAGGATGTGACAAATATATCAGGAGGAGTATCTGCAGAAACCCA 409
QY 1678 GTTGAGAACCATATCAATATATACACAAATCAGATACTTTCAGACCAAGCCATGGATTC 1737
Db 408 GTTGAGAACCATATCAATATATACACAAATCAGATACTTTCAGACCAAGCCATGGATTC 349
QY 1738 AACTCAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGGTTCAGAAAGAAATCT 1797
Db 348 AACTCAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGGTTCAGAAAGAAATCT 289
QY 1798 GCTTCATCTGAAAAAGCCAAAGGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1857
Db 288 GCTTCATCTGAAAAAGCCAAAGGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 229
QY 1858 ACCAATGAAATTAACACCAATCTGAGCTCAGTTCCTCAACAGAGGAGGAGGAGGAGGAGGAG 1917
Db 228 ACCAATGAAATTAACACCAATCTGAGCTCAGTTCCTCAACAGAGGAGGAGGAGGAGGAGGAG 169
QY 1918 GCTCATGAGAAACATCCACAGGAGTCTGCTCTTTCAGGACACAGATGACAGTGTATGAT 1977
Db 168 GCTCATGAGAAACATCCACAGGAGTCTGCTCTTTCAGGACACAGATGACAGTGTATGAT 109
QY 1978 GACCCAGTCTCTGATCCAGGTCAGAGTATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2037
Db 108 GACCCAGTCTCTGATCCAGGTCAGAGTATCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 49
QY 2038 GTTGCCCGTATTTCAGGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2085
Db 48 GTTGCCCGTATTTCAGGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1
RESULT 30
LOCUS BQ747286
DEFINITION BQ747286 634 bp mRNA linear EST 15-SEP-2000
ACCESSION BQ747286
VERSION BQ747286.1 GI:10161278
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 634)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9apbs-re@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: L1CM761 row: g column: 09
 High quality sequence stop: 634.
 Location/Qualifiers
 1. 634
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3929120"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 198 a 139 c 144 g 153 t
 ORIGIN

FEATURES
 source

Query Match 24.6%; Score 634; DB 12; Length 634;
 Best Local Similarity 100.0%; Pred. No. 2.9e-168;
 Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 773 CTCGTGTTACAGTGAAGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATAT 832
 Db 1 CTCGTGTTACAGTGAAGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATAT 60

QY 833 ATCTTTTTCACCCGAAAGATGATACAGCAGCAGAGTCTTAAACTCTCTCGGAGAGA 892
 Db 61 ATCTTTTTCACCCGAAAGATGATACAGCAGCAGAGTCTTAAACTCTCTCGGAGAGA 120

QY 893 GAAGAGAGAGTTGGGACACACACAGTTAAGCGTTTGAGACTTCGTGTTGTTGTCAG 952
 Db 121 GAAGAGAGAGTTGGGACACACACAGTTAAGCGTTTGAGACTTCGTGTTGTTGTCAG 180

QY 953 ATACTGGACCCAGACGAGCGGAGAGTGAACAGACGAGATGGAGCAGAGTCCCA 1012
 Db 181 ATACTGGACCCAGACGAGCGGAGAGTGAACAGACGAGATGGAGCAGAGTCCCA 240

QY 1013 ATGTGTCATGATGCAGAGATGTCGTATGTTATCAAGATGTTTGAAGAAGCAAGTG 1072
 Db 241 ATGTGTCATGATGCAGAGATGTCGTATGTTATCAAGATGTTTGAAGAAGCAAGTG 300

QY 1073 AGTTTGACAAAGCAATAGAGGACGAGAGATCTGCACCAGAGTGGAAACAAGTCAAT 1132
 Db 301 AGTTTGACAAAGCAATAGAGGACGAGAGATCTGCACCAGAGTGGAAACAAGTCAAT 360

QY 1133 CAGATATTTCAACTCTTCCTACAGTCCCATCAAGTCTGATTTGGAAGTGAAGTCACTG 1192
 Db 361 CAGATATTTCAACTCTTCCTACAGTCCCATCAAGTCTGATTTGGAAGTGAAGTCACTG 420

QY 1193 CAATGGAAGTATGATCTCCAGCTGAACAATTTCTTACGCCCTTCTACATCTCTACAATGT 1252
 Db 421 CAATGGAAGTATGATCTCCAGCTGAACAATTTCTTACGCCCTTCTACATCTCTACAATGT 480

QY 1253 CAGTCAAGGTCATCTCCACATCATCTCCACAGAAAGCCCTCATCTACCTCTTGTGAT 1312
 Db 481 CAGTCAAGGTCATCTCCACATCATCTCCACAGAAAGCCCTCATCTACCTCTTGTGAT 540

QY 1313 CTCTCCAGACAGTGAACAAGGAGTCTCTGTGAGGCATCTGGACACACACACATCATC 1372
 Db 541 CTCTCCAGACAGTGAACAAGGAGTCTCTGTGAGGCATCTGGACACACACACATCATC 600

QY 1373 AGTCTGTATACAAATATGAAAAGCTGAGCCCCAA 1406

Db 601 AGTCTGATACAAATATGAAAAGCTGAGCCCCAA 634

RESULT 31
 BM790042 635 bp mRNA linear EST 05-MAR-2002
 LOCUS K-EST0069804 S21SNU520 Homo sapiens cDNA clone S21SNU520-9-A04 5',
 DEFINITION mRNA sequence.
 ACCESSION BM790042
 VERSION BM790042.1 GI:19138274
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 635)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 9 row: A column: 04
 High quality sequence stop: 635.
 Location/Qualifiers
 1. 635
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="S21SNU520-9-A04"
 /clone_lib="S21SNU520"
 /sex="F"
 /tissue_type="Stomach"
 /cell_type="Floating aggregates"
 /lab_host="Top10F"
 /cell_line="SNU-520"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dt-selected mRNA by
 priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."
 BASE COUNT 218 a 148 c 144 g 125 t
 ORIGIN

Query Match 24.6%; Score 633.4; DB 14; Length 635;
 Best Local Similarity 99.8%; Pred. No. 4.3e-168;
 Matches 634; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1435 TTGCACGTACAGCAGAGGAGCAACTACACAGCAATAAACTCAACTTACAGATGAA 1494
 Db 1 TTGCACGTACAGCAGAGGAGCAACTACACAGCAATAAACTCAACTTACAGATGAA 60

QY 1495 TGGAGCAGTAGCATCAAGTCTTAGAGGAATTTGGAGCCATTCGAATCTGAGGGTCAG 1554
 Db 61 TGGAGCAGTAGCATCAAGTCTTAGAGGAATTTGGAGCCATTCGAATCTGAGGGTCAG 120

QY 1555 GAGGAATCTTTTGGTCCACAGAGCTCAGTGCACACAGAGGAGACAGTGAACAAAA 1614
 Db 121 GAGGAATCTTTTGGTCCACAGAGCTCAGTGCACACAGAGGAGACAGTGAACAAAA 180
 QY 1615 GCTCCTGAAGAATCATCAGAGGATGTGCACAAATATATCAGGAAGAGTATCTGCAGAAAC 1674
 Db 181 GCTCCTGAAGAATCATCAGAGGATGTGCACAAATATATCAGGAAGAGTATCTGCAGAAAC 240
 QY 1675 CCAGTTGAGAACCATATCAATATATACCAATCAGATAAGTTCACAGCCAGCCCAATTCGAT 1734
 Db 241 CCAGTTGAGAACCATATCAATATATACCAATCAGATAAGTTCACAGCCAGCCCAATTCGAT 300
 QY 1735 TCCAACTCAGAGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGTTCCAGAGAA 1794
 Db 301 TCCAACTCAGAGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGTTCCAGAGAA 360
 QY 1795 TCTGCTTCATCTGAAAAGCCAGGAACCAAGAACCTTCAGATCAGACTAGCAGTGAAGT 1854
 Db 361 TCTGCTTCATCTGAAAAGCCAGGAACCAAGAACCTTCAGATCAGACTAGCAGTGAAGT 420
 QY 1855 GCTACCAATGAAATAACACCAATCTTGAGCTCAGTTCACAAACAGAGCCACTGGGCT 1914
 Db 421 GCTACCAATGAAATAACACCAATCTTGAGCTCAGTTCACAAACAGAGCCACTGGGCT 480
 QY 1915 TCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGTAT 1974
 Db 481 TCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGTAT 540
 QY 1975 GATGACCCAGTCCCTGATCCAGGTCAGAGTATCAGAGCAGCTGATAGAGCTCT 2034
 Db 541 GATGACCCAGTCCCTGATCCAGGTCAGAGTATCAGAGCAGCTGATAGAGCTCT 600
 QY 2035 GCTGTTGCCCGTATTCAGGAGTCTTCAGAGGGAG 2069
 Db 601 GCTGTTGCCCGTATTCAGGAGTCTTCAGAGGGAG 635

RESULT 32
 BG719267 707 bp mRNA linear EST 08-MAY-2001
 LOCUS 602690335F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822641 5',
 DEFINITION mRNA sequence.
 ACCESSION BG719267.1 GI:13998454
 VERSION EST.
 KEYWORDS human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NIH-MGC <http://mgi.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLM0731 row: e column: 10
 High quality sequence stop: 704.
 Location/Qualifiers
 1..707
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4822641"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescript (modified)

FEATURES

source

RESULT 33
 BG534181

LOCUS 602553301F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:462816 5',
 DEFINITION mRNA sequence.

ACCESSION BG534181

VERSION BG534181.1 GI:13525721

KEYWORDS EST.

SOURCE human.

pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTT-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 210 a 139 c 151 g 207 t
 ORIGIN

Query Match 24.5%; Score 631.4; DB 12; Length 707;

Best Local Similarity 96.1%; Pred. No. 1.7e-167; Indels 1; Gaps 1;
 Matches 658; Conservative 0; Mismatches 26;

QY 210 AGATGACACCACTAGTAATTAATAGTATCCCTACAGCAGAAAGGTTTGTGACAACTTCG 269
 Db 23 AGAGCACTGAGAAATACCTTTTAAAGGAGAGGTTTCTCATCAGAGGTTTGTGACAACTTCG 82
 QY 270 TTCAGGGCAGCAGAGAAACATATTTAGTGCAAGTTCTTACTTGTACAAATGATAACA 329
 Db 83 TTCAGGGCAGCAGAGCAACATATTTAGTSCAAAGTTCTTACTTGTACAAATGATAACA 142
 QY 330 GATTGTATCTGCTGAGATGGAGTAATATTTATACCAAGTTGAGCAAGATGCAGA 389
 Db 143 GATTGTATCTGCTGAGATGGAGTAATATTTATACCAAGTTGAGCAAGATGCAGA 202
 QY 390 AACCAACAGACAATGCCAATTTACGTGTCTTATGGAATCTTATGAGATTTGACTGT 449
 Db 203 AACCAACAGACAATGCCAATTTACGTGTCTTATGGAATCTTATGAGATTTGACTGT 262
 QY 450 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGGAAGTGAAGTGTAGTGTTGA 509
 Db 263 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGGAAGTGAAGTGTAGTGTTGA 322
 QY 510 TACAGCATCAAACTAGCTGCACAAAAGAGATTGTAAAGATGATATTTTAACTG 569
 Db 323 TACAGCATCAAACTAGCTGCACAAAAGAGATTGTAAAGATGATATTTTAACTG 382
 QY 570 TCGAGTGTGCCAGTCTGTGCTATTTGCCCCCAATACCATATTTACCTTCTGCTGG 629
 Db 383 TCGAGTGTGCCAGTCTGTGCTATTTGCCCCCAATACCATATTTACCTTCTGCTGG 442
 QY 630 TTGTTCTGACAGCTCAGTACGAATATATGATCGGGAATCTGCGGCAAGAGCTACAGG 689
 Db 443 TTGTTCTGACAGCTCAGTACGAATATATGATCGGGAATCTGCGGCAAGAGCTACAGG 502
 QY 690 GAATTATGAGTTCGAGGAGCTACTGGAATGTTGCCCGTTTATTCCTTCCATCTTAA 749
 Db 503 GAATTATGAGTTCGAGGAGCTACTGGAATGTTGCCCGTTTATTCCTTCCATCTTAA 562
 QY 750 TAATAGTCTTCAGAGTGCACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCCT 809
 Db 563 TAATAGTCTTCAGAGTGCACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCCT 622
 QY 810 TAGTACTCTTCAGATTTACATATCTTTTACCCGGAAGATGATACAGCAGGAACT 869
 Db 623 TAGTACTCTTCAGATTTACATATCTTTTACCCGGAAGATGATACAGCAGGAACT 681
 QY 870 TAAACTCTTCTCGGGAAGAGAGA 894
 Db 682 TAAACTCTTCTCGGGAAGAGAGA 706

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 728)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 Plate: L1CMI1464 row: b column: 01
 High quality sequence stop: 681.
 FEATURES
 Location/Qualifiers
 1..728
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4662816"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: lung; Vector: pDMR-LIB (Clontech); Site: 1:
 SfiI (ggccgctcgccc); Site: 2: SfiI (ggccattggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CAGGCCATTATGCCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCCGCGGCGGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 221 a 153 c 175 g 179 t
 ORIGIN
 Query Match 24.4%; Score 629.8; DB 12; Length 728;
 Best Local Similarity 96.7%; Pred. No. 4.9e-167;
 Matches 707; Conservative 0; Mismatches 17; Indels 7; Gaps 6;
 QY 663 GCGAATGCTGGGCAAGAGCTACAGGGAATTATCAGGTCGAGGACTACTGGAATGTT 722
 Db 1
 QY 2 GGGAAATGCTGGGCAAGAGCTACAGGGAATTATCAGGTCGAGGACTACTGGAATGTT 61
 Db 1
 QY 723 TGCCCGTTTATTCCTCCCATCTTAATAAAGTCCTGCAGAGTGACATCTCTGTGTTA 782
 Db 62 TGCCCGTTTATTCCTCCCATCTTAATAAAGTCCTGCAGAGTGACATCTCTGTGTTA 121
 QY 783 CAGTGAAGATGCTCAAGAGATTCGTTAGTTACTCTTCAGATTCATATATCTTTTGA 842
 Db 122 CAGTGAAGATGCTCAAGAGATTCGTTAGTTACTCTTCAGATTCATATATCTTTTGA 181
 QY 843 CCGGAAAGATGATACAGCAGAGAACTTAAACCTCTTCGCGAAGAGAGAGAGAAGA 902
 Db 182 CCGGAAAGATGATACAGCAGAGAACTTAAACCTCTTCGCGAAGAGAGAGAGAAGA 241
 QY 903 GTTCCGACACACCACCTTAAGCGTTTGAGACTTCGTGGTGATTTGGTCAGATCTGGACC 962
 Db 242 G-TGCGAACAACACCTTAAGCG-TTGAGACTTCGTGGTGATTTGGTCAGATCTGGACC 298
 QY 963 CAGAGCAAGCGCGGAGAGTGAACGAGACGAGATGGAGAGCAGAGTCCCATGTGTCATT 1022
 Db 299 CAGAGCAAGCGCGGAGAGTGAACGAGACGAGATGGAGAGCAGAGTCCCATGTGTCATT 358
 QY 1023 GATGCGAGAGATGTCGATATGTTTATCAAGATGTTTGAAGAAGCAAGTGAGTTGACACA 1082
 Db 359 GATGCGAGAGATGTCGATATGTTTATCAAGATG-TTGAAGAAGCAAGTGAGTTGACACA 417
 QY 1083 AAGCAATAGGAGCAGAGATCTGACCCAGAGGTGGGAACAAGTCAATCAGATATTTTC 1142
 Db 418 AAGCAATAGGAGCAGAGATCTGACCCAGAGGTGGGAACAAGTCAATCAGATATTTTC 477

QY 1143 AACTTTCTTACGGTCCCATCAAGTCCCTGATTTGGGAAGTGAGTGAAACTGCAATGGAAGT 1202
 Db 478 AACTTTCTTACGGTCCCATCAAGTCCCTGATTTGGGAAGTGAGTGAAACTGCAATGGAAGT 537
 QY 1203 AGATACTCCAGCTGAACAATTTCTTCAGGCTTCTACATCTCTACATATGTCACCTCAGGC 1262
 Db 538 AGATACTCCAGCTGAACAATTTCTTCAGGCTTCTACATCTCTACATATGTCACCTCAGGC 597
 QY 1263 TCATTTCGACATCATCTCCACAGAAGCCCTCATTTCTACTCTCTTTCTCTCTCTCCAGA 1322
 Db 598 TCATTTCGA-ATCATCTCCACAGAAGCCCTCATTTCTACTCTCTCTCTCTCTCTCCAGA 656
 QY 1323 CAGTGAACAAGGCGAGCTCTGTTGAGG--CATCTGGACACACACATCATCATGTCGTGAT 1380
 Db 657 TAGTGAACAAGGCGAGCTCTGTTGAGGCGATCTGGAACACCAACCAAAACTTTCATCAGTCTGTT 716
 QY 1381 AACAAATATGA 1391
 Db 717 AACAAATATGA 727
 RESULT 34
 BI459472
 LOCUS
 DEFINITION 869 bp mRNA linear EST 21-AUG-2001
 603200466F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5266144 5',
 mRNA sequence.
 ACCESSION BI459472
 VERSION BI459472.1 GI:15250128
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 869)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM11671 row: d column: 17
 High quality sequence stop: 732.
 FEATURES
 Location/Qualifiers
 1..869
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5266144"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 257 a 163 c 214 g 235 t
 ORIGIN
 Query Match 24.4%; Score 628.6; DB 13; Length 869;
 Best Local Similarity 92.5%; Pred. No. 1.2e-166;
 Matches 750; Conservative 0; Mismatches 49; Indels 12; Gaps 8;

```

QY 1 ATGTCTGGGGTGGCTCTCTCCACACACCTGTGTGGGACGTGAGAAAAGTCTCCCTGGG 60
Db 38 ATGTCTGGGGTGGCTCTCTCCACACACCTGTGTGGGACGTGAGAAAAGTCTCCCTGGG 97
QY 61 CTGAGAGACCGCTCCCGGCTGGGAGTCTACCTGGGAGAGAGAAATTTATCCAAAGA 120
Db 98 CTGAGAGACCGCTCCCGGCTGGGAGTCTACCTGGGAGAGAGAAATTTATCCAAAGA 157
QY 121 TTAATACTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTATATAACAATCTGTGGAAAT 180
Db 158 TTAATACTTGAAGCAACCCCTTAATGTGCATGATGTTGTGTATATAACAATCTGTGGAAAT 217
QY 181 GACACTGGAGATAATATTTATCTGGCTCAGATGACACCAATTTAGTAATTCCT 240
Db 218 GACACTGGAGATAATATTTATCTGGCTCAGATGACACCAATTTAGTAATTCCT 277
QY 241 TACAGCAGAAAGGTTTGACAAACATTCGTTTCAGGACACCGAGCAACATATTTAGTGCA 300
Db 278 TACAGCAGAAAGGTTTGACAAACATTCGTTTCAGGACACCGAGCAACATATTTAGTGCA 337
QY 301 AAGTTCTTACCTTGTACAAATGATAAAGAGATTTGTCTCTCTGGAGATGGAGTAATA 360
Db 338 AAGTTCTTACCTTGTACAAATGATAAAGAGATTTGTCTCTCTGGAGATGGAGTAATA 397
QY 361 TTTTATACCAACGTTGACCAAGATGACAGAAACCAACAGCAATGCCAATTTACGTGTCAT 420
Db 398 TTTTATACCAACGTTGACCAAGATGACAGAAACCAACAGCAATGCCAATTTACGTGTCAT 457
QY 421 TATGGAACCTATGATGATATGACTCTACCAATGACCTTACACATTTCTCTCTGT 480
Db 458 TATGGAACCTATGATGATATGACTCTACCAATGACCTTACACATTTCTCTCTGT 517
QY 481 GGTGAAGATGGAACCT-GTTAGTGGTTTGTATACAGCATCAAAACTAGCTGCACAAAAGA 539
Db 518 GGTGAAGATGGAACCTGTTAGTGGTTTGTATACAGCATCAAAACTAGCTGCACAAAAGA 577
QY 540 AGATTGTAATTAATTAATCTGTCAGCTGCTGCCAGCTCTGTTGCTATTT 598
Db 578 AGATTGTAATTAATTAATCTGTCAGCTGCTGCCAGCTCTGTTGCTATTT 636
QY 599 GCCACCAATACCATTTACCTTGTGTTGTTCTGACAGCTCAGTAC-GAATATAT 657
Db 637 GCCGAC--AATACATATTTACCTTGTGTTGTTGTTCTGACAGCTCAGTACGGAATATAT 694
QY 658 GATCGGGGAATGCT--GGGCACAGAGCTACAGGGAATATTCAGGTCGAGGACTACTG 715
Db 695 GATCGGGGAATGCTGGGGGACAAAGAGTACAGGGAATATATTCAGGTCGAGGGAATATG 754
QY 716 GAATGGTT---GCCCGTTTATTCCTTCCATCTTAATAAATAGCTCCTGCAGAGTGACAT 772
Db 755 GAATGGTTGGCCGGTGTAGTTACTTGGCATTTTANTANTANTANTAGTCTGCGCAGGGA-AA 813
QY 773 CTCTGTGTACAGTGAAGATGTCAGAGAT 803
Db 814 CTCAGGGTTAAGTGAAGATGTCAGAGATAT 844

RESULT 35
LOCUS BE559905
DEFINITION 601345445F1 NIH_MGC_8 Homo sapiens cdna clone IMAGE:3678418 5',
rna sequence.
ACCESSION BE559905
VERSION BE559905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 762)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM355 row: i column: 11
High quality sequence stop: 690.

FEATURES
source

1..762
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3678418"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT
ORIGIN

Query Match 24.3%; Score 627.4; DB 10; Length 762;
Best Local Similarity 97.7%; Pred. No. 2.4e-166;
Matches 689; Conservative 0; Mismatches 11; Indels 5; Gaps 5;

```

QY 1378 GATAACAATAATGAAAAGCTGAGCCCCAAAACAGGACAGGTGAACCAAGTTTAAAGTTG 1437
Db 35 GATAACAATAATGAAAAGCTGAGCCCCAAAACCA-GGACAGGTGAACCAAGTTTAAAGTTG 93
QY 1438 CACTACACACAGAGGAACTACAGCAATATAAAGTGAACCTTTACAGATGAATGG 1497
Db 94 CACTACACACAGAGGAACTACAGCAATATAAAGTGAACCTTTACAGATGAATGG 153
QY 1498 AGCAGTATAGCATCAAGTTCTAGAGAAATTTGGAGCCATTTGCAAAATCTGAGGGTCAGGAG 1557
Db 154 AGCAGTATAGCATCAAGTTCTAGAGAAATTTGGAGCCATTTGCAAAATCTGAGGGTCAGGAG 213
QY 1558 GAATCTTTCTGTCACAGAGCTCAGTGCACCAACAGAGGAGAGACAGTGAACCAAAAGCT 1617
Db 214 GAATCTTTCTGTCACAGAGCTCAGTGCACCAACAGAGGAGAGACAGTGAACCAAAAGCT 273
QY 1618 CCTGAAGATCATCAGAGGATGTGACAAATATCAGGAAGAGTATCTGCAGAAACCCA 1677
Db 274 CCTGAAGATCATCAGAGGATGTGACAAATATCAGGAAGAGTATCTGCAGAAACCCA 333
QY 1678 GTTGAGAACCATATCAATATATACAAATCAGATAGTTTCACAGCCAGCCATTTGGATTCC 1737
Db 334 GTTGAGAACCATATCAATATATACAAATCAGATAGTTTCACAGCCAGCCATTTGGATTCC 393
QY 1738 AACTCAGGAGAAAGAAATGACCTCAATCTTGTGCTGTGGGTTTCCAGAAAGAACTCT 1797
Db 394 AACTCAGGAGAAAGAAATGACCTCAATCTTGTGCTGTGGGTTTCCAGAGAAATCTCT 453
QY 1798 GCTTCTATCTGAAAAAGCCAGGAACCAAGAACTTTCAGATCAGACTAGCAGTGTGTGTGT 1857
Db 454 GCTTCTATCTGAAAAAGCCAGGAACCAAGAACTTTCAGATCAGACTAGCAGTGTGTGTGT 513
QY 1858 ACCAATGAAAAATACCAATCTCTGAGCCCTCAGTTCCAAACAGAAAGCCACTGGCCCTCA 1917
Db 514 ACCAATGAAAAATACCAATCTCTGAGCCCTCAGTTCCAAACAGAAAGCCACTGGCCCTCA 573
QY 1918 GCTCATGAAGAAACATCCACCGAGGACTCTGCTCTTTCAGGACACAGATG-ACAGTGATGA 1976
Db 574 GCTCATGAAGAAACATCCACCGAGGACTCTGCTCTTTCAGGACACACATGTACAGTATGA 633

```


); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0.5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 222 a 144 c 156 g 211 t
ORIGIN

Query Match 24.2%; Score 624.4; DB 13; Length 733;
Best Local Similarity 95.9%; Pred. No. 1.7e-165;
Matches 673; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

```

QY 220 AAATTAGTAACTTACGATCCCTTACAGCAGAAAGGTTTGGACAACTTCGTTACGGGCAC 279
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 32 AAATACCTTTTAAAGGAGAGGTTTCTCATCAGAGGTTTGGACAACTTCGTTACGGGCAC 91
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 280 CGAGCAACATATTTAGTGCAGAAAGTCTTACCTTGTACAAATGATAACACAGATTGTATCC 339
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 92 CGAGCAACATATTTAGTGCAGAAAGTCTTACCTTGTACAAATGATAACACAGATTGTATCC 151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 340 TGTCTGGAGATGAGATTAATTTATACCAAGCTTGAGCAAGATGCAAGAACCAACAGAGA 399
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 152 TGTCTGGAGATGAGATTAATTTATACCAAGCTTGAGCAAGATGCAAGAACCAACAGAGA 211
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 400 CAATGCCAATTTACGTGTCATTTATGGAAGTCTTATGAGATTTATGACGTACCCCAATGAC 459
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 212 CAATGCCAATTTACGTGTCATTTATGGAAGTCTTATGAGATTTATGACGTACCCCAATGAC 271
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 460 CCTTACACTTTTCTCTCTGTTGGTGAAGATGGAAGTCTTGTAGTGGTTTGTATACAGCATC 519
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 272 CCTTACACTTTTCTCTCTGTTGGTGAAGATGGAAGTCTTGTAGTGGTTTGTATACAGCATC 331
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 520 AAACTAGCTGCACAAAAGAGATTGAAAGATGATATTTTAATTAACCTGTCGAGCTGCT 579
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 332 AAACTAGCTGCACAAAAGAGATTGAAAGATGATATTTTAATTAACCTGTCGAGCTGCT 391
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 580 GCCAGCTGTTGCTTAATTTGCCACCAATACCATATTAACCTGCTGTTGTTGTTCTGAC 639
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 392 GCCAGCTGTTGCTTAATTTGCCACCAATACCATATTAACCTGCTGTTGTTGTTCTGAC 451
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 640 AGCTCAGTACGAATATATGTCGGGAATGCTGGGCACAAAGAGCTACAGGGAATTTATGCA 699
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 452 AGCTCAGTACGAATATATGTCGGGAATGCTGGGCACAAAGAGCTACAGGGAATTTATGCA 511
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 700 GGTCTGAGGAGTACTGGAATGTTGCCCGTTTATTCCTTCCCATCTTAATAATAAGTCC 759
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 512 GGTCTGAGGAGTACTGGAATGTTGCCCGTTTATTCCTTCCCATCTTAATAATAAGTCC 571
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 760 TGCAGAGTACATCTCTGTGTTACAGTGAAGATGTTCAAGAGATTCTCGTTAGTTACTCT 819
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 572 TGCAGAGTACATCTCTGTGTTACAGTGAAGATGTTCAAGAGATTCTCGTTAGTTACTCT 631
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 820 TCAGATTACATATATCTTTTTCACCGG-AAAGATGATACAGCAGCAGAAC-TTAAACATC 877
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 632 TCAGATTACATATATCTTTTTCACCGGAAAGATGATACAGCAGCAGAACCTTAAACATC 691
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 878 CTTCTCGGGAAGAG-AGAGAGAAGAGTTTGCACACCAACCA 918
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 692 CTTCTCGGGAAGAGAGAGAGAGAGTTGCGACACCA 733
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 38
BM926145
LOCUS BM926145
DEFINITION AGENCOURT_6600775 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5764509
5', mRNA sequence.
ACCESSION BM926145
VERSION BM926145.1 GI:19376524
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 1199)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLML at:
<http://image.llnl.gov>

Plate: LLML2818 row: a column: 22
High quality sequence start: 28
High quality sequence stop: 546.

FEATURES

Location/Qualifiers
1..1199
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5764509"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."

BASE COUNT 359 a 290 c 251 g 292 t 7 others
ORIGIN

Query Match 24.2%; Score 624; DB 14; Length 1199;

Best Local Similarity 96.4%; Pred. No. 2.8e-165;

Matches 692; Conservative 0; Mismatches 20; Indels 6; Gaps 5;

```

QY 44 GGAAGAAGTCCCTCGGCTGGAGAGCCCGTCCCGCTCGGAGTCCGTACCTGGGAAGAA 103
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 5 GCAAAAGGGGGCTCGGCTGGAGGA-ATGCCCGGCTCGGAGTCCGTACCTGGGAAGAA 63
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 104 GAGATTTATCAAGATTAAACTTGAGCAACCTTAATGTCATGATGTTGTGTTA 163
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 64 GAGAATTTATCAAGATTATATCTTGAAGCAACCCCTTAATGTCATGATGTTGTGTTA 123
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 164 ATACAATCTGTTGGAGTACACT-GGAGAAATATATTTATCTGGCTCAGATGACACAAA 222
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 124 ATACAATCTGTTGGAGTACACTGGAGAAATATTTATCTGGCTCAGATGACACAAA 183
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 223 TTAGTAATTTAGTAATCTTTACAGCAAGAAAGTTTGGACAACTTCGTTACAGGCAACGA 282
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 184 TTAGTAATTTAGTAATCTTTACAGCAAGAAAGTTTGGACAACTTCGTTACAGGCAACGA 243
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 283 GCAACATATTTAGTGCAGAAAGTCTTACCTTGTACAAATGATAACAGATTGTATCTCTC 342
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 244 GCAACATATTTAGTGCAGAAAGTCTTACCTTGTACAAATGATAACAGATTGTATCTCTC 303
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 343 TCTGGAGATGGAGTAATATTTATACCAAGCTTGACAAAGATGCAGAAACCAACAGACAA 402
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 304 TCTGGAGATGGAGTAATATTTATACCAAGCTTGACAAAGATGCAGAAACCAACAGACAA 363
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 403 TGCCTAATTTACGTGTCATTTATGGAATCTACTTATGAGATTATGACTTACCCAATGACCT 462
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 364 TGCCTAATTTACGTGTCATTTATGGAATCTACTTATGAGATTATGACTTACCCAATGACCT 423
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 463 TACACTTTTCTCTCTGTTGGTGAAGATGGAAGTGTAGTGGTTTGTATACACCATCAAA 522
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 424 TACACTTTTCTCTCTGTTGGTGAAGATGGAAGTGTAGTGGTTTGTATACACCATCAAA 483
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```


found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L1CM849 row: o column: 22
High quality sequence stop: 587.
Location/Qualifiers
1. 912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3996909"
/clone_lib="NIH_MGC_53"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgctcgcc); Site_2: SfII (ggccattggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATTCTAGAGCGGCGGCCGACATG-dT(30)BN-3'
sequence: 5'-ATTCTAGAGCGGCGGCCGACATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT	294 a	216 c	228 g	174 t
ORIGIN				
Query Match	24.0%	Score 618.4;	DB 12;	Length 912;
Best Local Similarity	97.1%	Pred. No. 9.3e-164;		
Matches	662;	Conservative	0;	Mismatches 16; Indels 4; Gaps 3;
Qy	1575	GAGCTCAGTGAACACACAGAGGAGAGAGTCTCGAGAAACCCAGTTGAGAACCATATCAA	1634	
Db	1	GAGCTCAGTGAACACACAGAGGAGAGTCTCGAGAAACCCAGTTGAGAACCATATCAA	60	
Qy	1635	GGATGTGACAAATATCAGGAGGAGTCTCGAGAAACCCAGTTGAGAACCATATCAA	1694	
Db	61	GGATGTGACAAATATCAGGAGGAGTCTCGAGAAACCCAGTTGAGAACCATATCAA	120	
Qy	1695	TATAACACAATCAGATAAGTTACAGCCAAAGCCATTGGATTCCAACTCAGGAGAGAGAA	1754	
Db	121	TATAACACAATCAGATAAGTTACAGCCAAAGCCATTGGATTCCAACTCAGGAGAGAGAA	180	
Qy	1755	TCACCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCTCATCTGAAAAAGC	1814	
Db	181	TCACCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCTCATCTGAAAAAGC	240	
Qy	1815	CAAGGAACAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAATAACAC	1874	
Db	241	CAAGGAACAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAATAACAC	300	
Qy	1875	CAATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAACATC	1934	
Db	301	CAATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAACATC	360	
Qy	1935	CACAGGGACTCTGCTCTTCAGGACACATGATGATGATGATGATGATGATGATGATGATG	1994	
Db	361	CACAGGGACTCTGCTCTTCAGGACACATGATGATGATGATGATGATGATGATGATGATG	420	
Qy	1995	AGGTGCAAGGTATCGAGCAGGACCTGGTGTATAGACGCTCTGCTGTGTCGCCGTTATCAGGA	2054	
Db	421	AGGTGCAAGGTATCGAGCAGGACCTGGTGTATAGACGCTCTGCTGTGTCGCCGTTATCAGGA	480	
Qy	2055	GTTCCTCAGACGGAGAAAGAGAGAAAGAAATGGAGAAATGGATACCTTGGACATTAG	2114	
Db	481	GTTCCTCAGACGGAGAAAGAGAGAAAGAAATGGAGAAATGGATACCTTGGACATTAG	540	
Qy	2115	AAGGCGCTAGTAAATGGTTTATAAAGCCATCCCACTCCAGGACCAATGATAAAGA	2174	
Db	541	AAGGCGCTAGTAAATGGTTTATAAAGCCATCCCACTCCAGGACCAATGATAAAGA	596	
Qy	2175	AGCCAAATTTCTGGGGTCTAACTTTTAAATAGTGTGTTCTGACTGTGGCCACATTTTCAT	2234	
Db	597	AGCCAAATTTCTGGGGTCTAACTTTTAAATAGTGTGTTCTGACTGTGGCCACATTTTCAT	656	

QY 2235 CTGGGATCGGCACACTGCTGAG 2256
||||||| ||||| ||
Db 657 CTGGGATCGGAACACTGTGAAG 678

Search completed: March 26, 2003, 03:07:25
Job time : 4101 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2003, 20:23:58 ; Search time 104 seconds

(without alignments)

7607.943 Million cell updates/sec

Title: US-09-781-693A-1

Perfect score: 2580

Sequence: 1 atgtctgggggtgctcta.....atgaaatgagatgagaa 2580

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	59	2.3	7218	1	US-08-232-463-14
2	39	1.5	429	4	US-08-847-065-20
3	39	1.5	2462	4	US-08-847-065-24
4	39	1.5	3023	4	US-09-308-022-4
5	38.4	1.5	2085	1	US-08-072-070-1
6	38.4	1.5	2085	1	US-08-465-746-1
7	38.4	1.5	2085	1	US-08-214-164-1
8	38.4	1.5	2085	1	US-08-469-434-1
9	38.4	1.5	2085	1	US-08-214-222-1
10	38.4	1.5	2085	2	US-08-467-852A-1
11	38.4	1.5	2085	2	US-08-468-718-1
12	38.4	1.5	2085	2	US-08-246-636-1
13	38.4	1.5	2085	2	US-08-247-491A-1
14	38.4	1.5	2085	2	US-08-319-795-1
15	38.4	1.5	2085	2	US-08-468-985-1
16	38.4	1.5	2086	3	US-08-312-949-1
17	38.4	1.5	2086	3	US-08-446-201-2
18	37.4	1.4	1360	3	US-08-961-083-37
19	37.4	1.4	12665	4	US-08-961-527-134
20	35.8	1.4	289	4	US-09-007-005-17
21	35.8	1.4	289	4	US-09-244-796-17
22	35.6	1.4	1695	4	US-09-308-022-5
23	35.4	1.4	598	4	US-09-385-983-267
C 24	35.4	1.4	1188	3	US-08-048-700-1
C 25	35.4	1.4	1188	3	US-08-839-711-5
C 26	34.6	1.3	405	2	US-08-299-074A-1
C 27	34.6	1.3	405	4	US-09-399-773-1

C 28	34.6	1.3	1815	2	US-08-606-288-6	Sequence 6, Appli
C 29	34.6	1.3	1815	3	US-09-347-483-6	Sequence 6, Appli
30	34.4	1.3	248	4	US-09-007-005-32	Sequence 32, Appli
31	34.4	1.3	248	4	US-09-244-796-32	Sequence 32, Appli
32	34.4	1.3	277	4	US-09-007-005-3	Sequence 3, Appli
33	34.4	1.3	277	4	US-09-244-796-3	Sequence 3, Appli
34	34.4	1.3	913	1	US-08-217-327-3	Sequence 2, Appli
35	34.4	1.3	913	1	US-07-885-970A-3	Sequence 2, Appli
36	34.4	1.3	913	1	US-08-298-687A-3	Sequence 2, Appli
37	34.4	1.3	913	1	US-08-530-797-2	Sequence 2, Appli
38	34.4	1.3	913	1	US-08-298-829-3	Sequence 2, Appli
39	34.4	1.3	913	2	US-08-787-335-2	Sequence 2, Appli
40	34.4	1.3	1984	1	US-07-885-970A-25	Sequence 25, Appli
41	34.4	1.3	1985	1	US-08-298-829-25	Sequence 25, Appli
42	34.4	1.3	7100	4	US-09-308-375-1	Sequence 1, Appli
43	33.6	1.3	825	4	US-09-134-001C-2307	Sequence 1084, Ap
44	33.4	1.3	924	4	US-09-461-697-205	Sequence 205, App
45	33.4	1.3	282	4	US-09-461-697-193	Sequence 193, App
46	33.2	1.3	306	4	US-09-461-697-191	Sequence 191, App
47	33.2	1.3	696	4	US-09-461-697-189	Sequence 189, App
48	33.2	1.3	699	4	US-09-461-697-187	Sequence 187, App
49	33.2	1.3	717	4	US-09-123-912-113	Sequence 185, App
50	33.2	1.3	774	4	US-09-643-597-113	Sequence 113, App
51	33.2	1.3	819	4	US-09-060-410-3	Sequence 184, App
52	33.2	1.3	957	4	US-08-929-329-1	Sequence 3, Appli
C 53	33.2	1.3	957	4	US-08-078-080-3	Sequence 3, Appli
C 54	33.2	1.3	1669	4	US-08-930-055A-1	Sequence 1, Appli
C 55	33.2	1.3	4296	4	US-08-930-055A-2	Sequence 2, Appli
C 56	33.2	1.3	5433	3	US-08-899-241-1	Sequence 1, Appli
C 57	33	1.3	10607	1	US-09-347-803-11	Sequence 11, Appli
C 58	33	1.3	10607	1	US-09-115-704-1	Sequence 1, Appli
C 59	32.8	1.3	4617	4	US-08-961-527-137	Sequence 137, App
60	32.8	1.3	5109	4	US-07-757-022B-37	Sequence 37, App
61	32.8	1.3	5109	4	US-07-757-022B-31	Sequence 31, Appl
62	32.6	1.3	1683	4	US-09-134-001C-1565	Sequence 1565, Ap
63	32.6	1.3	2652	4	US-09-009-217-11	Sequence 11, Appl
C 64	32.6	1.3	12666	4	US-09-360-186-1	Sequence 1, Appli
C 65	32.4	1.3	444	4	US-09-328-111-9	Sequence 9, Appli
66	32.4	1.3	567	4	US-08-937-271-5	Sequence 5, Appli
67	32.4	1.3	1665	4	US-08-582-126-54	Sequence 54, Appl
C 68	32.4	1.3	13865	3	US-08-403-866-11	Sequence 11, Appl
C 69	32.4	1.3	13865	3	US-08-781-891-207	Sequence 207, App
C 70	32.4	1.3	31491	4	US-09-397-992A-3	Sequence 3, Appli
C 71	32.2	1.2	645	4	US-09-284-782-21	Sequence 21, Appl
C 72	32.2	1.2	822	3	US-09-134-001C-747	Sequence 747, App
C 73	32.2	1.2	5893	1	US-09-072-596-298	Sequence 298, App
74	32.2	1.2	12720	1	US-08-212-188-1	Sequence 1, Appli
C 75	32.2	1.2	29604	3	US-08-970-725-1	Sequence 1, Appli
C 76	32	1.2	624	4	PCT-US96-0520A-637	Sequence 637, App
C 77	32	1.2	624	4	US-08-940-751-4	Sequence 4, Appli
C 78	31.8	1.2	497	3	US-08-291-368-1	Sequence 1, Appli
C 79	31.8	1.2	606	4	US-08-962-190-1	Sequence 1, Appli
C 80	31.8	1.2	921	4	PCT-US95-10310-1	Sequence 1, Appli
C 81	31.8	1.2	1062	4	US-09-318-794A-4	Sequence 4, Appli
C 82	31.8	1.2	2799	3	US-08-293-728-1	Sequence 1, Appli
C 83	31.8	1.2	2799	3	US-09-421-868-1	Sequence 1, Appli
C 84	31.8	1.2	2799	5	US-08-682-517-13	Sequence 13, Appl
C 85	31.8	1.2	2799	5	US-08-682-517-14	Sequence 14, Appl
C 86	31.8	1.2	3522	1	US-08-682-517-7	Sequence 7, Appli
C 87	31.8	1.2	3522	1	US-08-682-517-8	Sequence 8, Appli
C 88	31.8	1.2	3522	2	US-09-080-897-3	Sequence 3, Appli
C 89	31.8	1.2	3522	6	US-08-899-595-2	Sequence 2, Appli
C 90	31.8	1.2	2952	4		
C 91	31.6	1.2	3498	3		
C 92	31.6	1.2	3498	4		
C 93	31.6	1.2	3498	4		
C 94	31.6	1.2	3666	2		
C 95	31.6	1.2	3666	2		
C 96	31.6	1.2	4197	2		
C 97	31.6	1.2	4197	2		
C 98	31.6	1.2	4378	2		
C 99	31.6	1.2	4378	2		
100	31.6	1.2	4399	3		

Patent No. 5164180

; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: S. pneumoniae
 US-08-847-065-20

Query Match 1.5%; Score 39; DB 4; Length 429;
 Best Local Similarity 45.1%; Pred. No. 0.053;
 Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCCTGAAGAACATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGA 1670
 DB 1 AAAAAGGTAGCAGAAGCTGAGAAGGTTGAAGAAGCTGAGAAAAAGCCAAAGGATCA 60
 QY 1671 AAACCCAGTTGAGAACCATATCAATATACACAAATCAGATAAGTTACAGGCCAAGCCATT 1730
 DB 61 AAAAGAAGAAGATCCCGTAACTACCCAACTACTTACAAAACGCTTGACCTTGAAT 120
 QY 1731 GGATTCCTCACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTGTGGGTTCCAGA 1790
 DB 121 TGCTGAGTCCGATGGAAGTTAAGAAGCGGAGCTTGAACTAGTAAAGAGGAAGCTAA 180
 QY 1791 AGAATCTGCTTCATCTGAAAAGCCCAAGAACCCAGAACTTCAGATCAGACTAGCACTGA 1850
 DB 181 GGAACCTCGAGACGAGGAGGAAAAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAAGC 240
 QY 1851 GAGTGCTACCAATGAAATAACACCAATCCTGAGCTCAGTTCCAAACAGAGCCACTGG 1910
 DB 241 TGAGGCTACAGGTTAGAAACATCAAGACAGATCGTAAAAAGCAGAGAAGAGCTAA 300
 QY 1911 GCCTTCAGCTCATGAAGAA 1929
 DB 301 ACGAAAGCAGCAGAAGAA 319

RESULT 3

US-08-847-065-24

; Sequence 24, Application US/08847065
 ; Patent No. 6245335
 ; GENERAL INFORMATION:
 ; APPLICANT: Masure, H. Robert
 ; APPLICANT: Rosenow, Carsten I.
 ; APPLICANT: Tuomanen, Elaine
 ; APPLICANT: Wizemann, Theresa M.
 ; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
 ; TITLE OF INVENTION: ANTI-PNEUMOCOCCAL VACCINES
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David A. Jackson, Esq.
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/847,065
 ; FILING DATE:

; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-158
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; INFORMATION FOR SEQ ID NO: 24:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2462 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; US-08-847-065-24

Query Match 1.5%; Score 39; DB 4; Length 2462;
 Best Local Similarity 45.1%; Pred. No. 0.16;
 Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCCTGAAGAACATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGA 1670
 DB 768 AAAAAGGTAGCAGAAGCTGAGAAGGTTGAAGAAGCTGAGAAAAAGCCAAAGGATCA 827
 QY 1671 AAACCCAGTTGAGAACCATATCAATATACACAAATCAGATAAGTTACAGGCCAAGCCATT 1730
 DB 828 AAAAGAAGAAGATCCCGTAACTACCCAACTACTTACAAAACGCTTGACCTTGAAT 887
 QY 1731 GGATTCCTCACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTGTGGGTTCCAGA 1790
 DB 888 TGCTGAGTCCGATGGAAGTTAAGAAGCGGAGCTTGAACTAGTAAAGAGGAAGCTAA 947
 QY 1791 AGAATCTGCTTCATCTGAAAAGCCCAAGAACCCAGAACTTCAGATCAGACTAGCACTGA 1850
 DB 948 GGAACCTCGAGACGAGGAGGAAAAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAAGC 1007
 QY 1851 GAGTGCTACCAATGAAATAACACCAATCCTGAGCTCAGTTCCAAACAGAGCCACTGG 1910
 DB 1008 TGAGGCTACAGGTTAGAAACATCAAGACAGATCGTAAAAAGCAGAGAAGAGCTAA 1067
 QY 1911 GCCTTCAGCTCATGAAGAA 1929
 DB 1068 ACGAAAGCAGCAGAAGAA 1086

RESULT 4

US-09-308-022-4

; Sequence 4, Application US/09308022
 ; Patent No. 6291654
 ; GENERAL INFORMATION:
 ; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
 ; TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
 ; TITLE OF INVENTION: PNEUMONIAE
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
 ; STREET: 119 No. 6291654th Fourth Street, Suite 203
 ; CITY: Minneapolis
 ; STATE: Minnesota
 ; COUNTRY: USA
 ; ZIP: 55401

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/308,022
 ; FILING DATE:

; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US97/20586
 ; FILING DATE: 12-NOV-1997
 ; APPLICATION NUMBER: 60/029,444
 ; FILING DATE: 12-NOV-1996
 ; APPLICATION NUMBER: 60/038,086
 ; FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
CLASSIFICATION: 60/062,473
PRIORITY APPLICATION DATA:
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-308-022-4

Query Match 1.5%; Score 39; DB 4; Length 3023;
Best Local Similarity 45.1%; Pred. No. 0.19;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGA 1670
DB 1282 AAAAAAGGTAGCAGAGCTGAGAAGGTTGAAGAAGCTGAGAAAAAAGCCAGGATCA 1341
QY 1671 AAACCCAGTTGAGAACCATATCAATATAACACATCAGATCAAGTTTCACAGCCAAAGCCATT 1730
DB 1342 AAAAGAGAAGATCGCCGTAACTACCCAAACCAATCTTACAAAAGCTTGACCTTGAAT 1401
QY 1731 GGATTCCAACCTCAGGAGAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTCCAGA 1790
DB 1402 TGCTGAGTCCGATGGAAGTTAAAGAGCGGAGCTTCACTAGTAAAAAGAGGAGCTAA 1461
QY 1791 AGAATCTCTTCAATCTGAAAAGCGAAGAACCCAGAACTTCAGATCAGATAGCAGTGA 1850
DB 1462 GGAACCTCAGAGCAGGAGAAAATTAAGCAAGCAAAAGGTTGAGAGTAAAGGCTAA 1521
QY 1851 GAGTGCTACCAATGAAATACACCAATCTCAGCTCAGCTTCCAAACAGAGAGCCACTGG 1910
DB 1522 TGAGCTACAGGTAGAAACATCAACACAGATCGTAAAGAGCAGAGAGAGCTAA 1581
QY 1911 GCCTTCAGCTCATGAAGAA 1929
DB 1582 ACGAAAAGCAGCAGAGAA 1600

RESULT 5
US-08-072-070-1
Sequence 1, Application US/08072070
Patent No. 5476929
GENERAL INFORMATION:
APPLICANT: Briles, David E
APPLICANT: Yother, Janet L
APPLICANT: McDaniel, Larry S
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: Suite 1203, 2001 Jefferson Davis Highway
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/072,070
FILING DATE: 19930603
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/835,698
FILING DATE: 12-FEB-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/656,773
FILING DATE: 15-FEB-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 521-0378
TELEX: LUKPAT WASHINGTON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: Rxi
IMMEDIATE SOURCE:
CLONE: JY2008
FEATURE:
NAME/KEY: intron
LOCATION: 1..2085
FEATURE:
NAME/KEY: CDS
LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
LOCATION: ..2025, 2029..2031, 2035..2085)
US-08-072-070-1

Query Match 1.5%; Score 38.4; DB 1; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAAGCTCCGGAAGATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCT 1665
DB 397 GAAAAGCAGCGCTCTGGAAGAGATGGTAAGCGAGTGGCAGCAGTTCACAAAGCGTATCTA 456
QY 1666 GCAGAAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCAA 1725
DB 457 GCCTATCAACAGCTACAGACAAACGCCGAAAGAGCGCAGCAGATAGATGATGAA 516
QY 1726 CCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGTGCTTGTGGGGTT 1785
DB 517 GCTAAGAAACGCGAAGAGAGGCAAAACTAAATTTAATCTGTTTCGAGCAATGGTAGTT 576
QY 1786 CCAGAAAGATCTGTTTCATCTGAAAAGCCAGGAAACCAAGCAACTTCAGATCAGACTAGC 1845
DB 577 CCTGAGCCAGAGCAGTGTGGCTGAGACTAGAAAAAATCAGAAAGAGCTAAACAAAAGCA 636
QY 1846 ACTGAGAGTCTACCAATGAAAATAACACCAA 1877
DB 637 CCAGAACTTACTAAAAAACTAGAAAGCTAA 668

RESULT 6
US-08-465-746-1
Sequence 1, Application US/08465746
Patent No. 5679768
GENERAL INFORMATION:
APPLICANT: Briles, David E
APPLICANT: Yother, Janet L
APPLICANT: McDaniel, Larry S
TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
TITLE OF INVENTION: PROTEIN A
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

```

1  RESULT 7
2  US-08-214-164-1
3  ; Sequence 1, Application US/08214164
4  ; Patent No. 5728387
5  ; GENERAL INFORMATION:
6  ; APPLICANT: BRILES, DAVID E.
7  ; APPLICANT: YOTHER, JANET L.
8  ; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
9  ; NUMBER OF SEQUENCES: 3
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Shoemaker and Mattare, Ltd
12 ; STREET: Suite 1203, 2001 Jefferson Davis Highway
13 ; CITY: Arlington
14 ; STATE: Virginia
15 ; COUNTRY: U.S.A.
16 ; ZIP: 22202-0286
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: IBM PC compatible
20 ; OPERATING SYSTEM: PC-DOS/MS-DOS
21 ; SOFTWARE: Patent In Release #1.0, Version #1.25
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/08/214,164
24 ; FILING DATE: 17-MAR-1994
25 ; CLASSIFICATION: 424
26 ; PRIOR APPLICATION DATA:
27 ; APPLICATION NUMBER: US 07/656,773
28 ; FILING DATE: 15-FEB-1991
29 ; CLASSIFICATION: 424
30 ; ATTORNEY/AGENT INFORMATION:
31 ; NAME: Berkstresser, Jerry W.
32 ; REGISTRATION NUMBER: 22,651
33 ; REFERENCE/DOCKET NUMBER: 6102-137
34 ; TELECOMMUNICATION INFORMATION:
35 ; TELEPHONE: (703) 415-0810
36 ; TELEFAX: (703) 521-0813
37 ; TELEX: LUKPAT WASHINGTON
38 ; INFORMATION FOR SEQ ID NO: 1:
39 ; SEQUENCE CHARACTERISTICS:
40 ; LENGTH: 2085 base pairs
41 ; TYPE: nucleic acid
42 ; STRANDEDNESS: double
43 ; TOPOLOGY: linear
44 ; MOLECULE TYPE: DNA (genomic)
45 ; FEATURE:
46 ; NAME/KEY: intron
47 ; LOCATION: 1..1983
48 ; FEATURE:
49 ; NAME/KEY: CDS
50 ; LOCATION: 127..1983
51 ; US-08-214-164-1

```

Query Match	1.5%	Score 38.4	DB 1	Length 2085
Best Local Similarity	46.3%	Pred. No. 0.23		
Matches 126	Conservative 0	Mismatches 146	Indels 0	Gaps 0
QY 1606	GAACAAAGCTCCTGAAGAAATCATCAGAGATGTGACAAAATATCAGSAAGGATATCT	1665		
DB				
QY 397	GA AAAAGCAGCGTCTGGAAGAGATGGATAAGCGCAGTGGCAGCGATTCACACAGCGTATCTA	456		
DB				
QY 1666	GCAGAAAACCCAGTTGACAGCCATATCAATATACACAATCAGATTAAGTTCACAGCCAAAG	1725		
DB				
QY 457	GCCTATCAACAAAGCTACAGACAAAACCGCCAAAAGACGCGACAGATAAGATGATGATGAA	516		
DB				
QY 1726	CCATTTGGATTCCCAACTCAGGAGAAAGAAATCACCTCAATCTTGATTCGCTTTGTGGGGTT	1785		
DB				
QY 517	GCTAAGAAACGCGAAGAAGAGCGCAAAACTAAATTTAATCTGTTTCGAGCAATGGTAGTT	576		
DB				
QY 1786	CCAGAAGNAATCTGCTTCATCTGTA AAAAGCCCAAGGAACCGAGAAATCTCAGATCAGACTAGC	1845		
DB				
QY 577	CCTGACGCCAGCAGCTGGCTGGACATTAAGAAAAAATTCAGAAAGAGCTTAAACAAAAAGCA	636		
DB				

IMMEDIATE SOURCE:


```

; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; NAME/KEY: CDS
; LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
; LOCATION: ..2025, 2029..2031, 2035..2085)
; US-08-468-718-1

Query Match 1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCTGGAAGATCATCAGAGGATGACAAATATCAGGAGGAGTATCT 1665
DB 397 GAAACAAAGCTCTGGAAGATCATCAGAGGATGACAAATATCAGGAGGAGTATCTA 456
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAGACGAGATGATGATGATGAA 516
QY 1726 CCATTGGATTCCAACTCAGGAGAAAGATGACCTCAATCTTGTCGCTCTTTGGGGTT 1785
DB 517 GCTAAGAAACGCGAGAGAGGCAAAACTAAATTAATCTGTTGAGCAATGTTAGTT 576
QY 1786 CCAGAGATCTGCTTCATCTGAAAGCCCAAGGACCAAGAACTTCAGATCAGACTAGC 1845
DB 577 CCGAGCAGAGCAGCTGGCTGAGACTAAGAAATAATCAGAGAAAGCTTAACAAAGCA 636
QY 1846 ACTGAGAGTGTACCAATGAAAATAACACCA 1877
DB 637 CCAGAACTTACTAAAAAAGCTAAGAGAGCTAA 668

RESULT 12
US-08-246-636-1
; Sequence 1, Application US/08246636
; Patent No. 5965141
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Yother, Janet L.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Wu, Hong-Yin
; TITLE OF INVENTION: EPTOTIC REGIONS OF PNEUMOCOCCAL SURFACE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246, 636
; FILING DATE: 20-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656, 773
; FILING DATE: 15-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835, 698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048, 896
; FILING DATE: 20-APR-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY4313
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; NAME/KEY: CDS
; LOCATION: join(127..1984)
; US-08-246-636-1

Query Match 1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCTGGAAGATCATCAGAGGATGACAAATATCAGGAGGAGTATCT 1665
DB 397 GAAACAAAGCTCTGGAAGATCATCAGAGGATGACAAATATCAGGAGGAGTATCTA 456
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAGACGAGATGATGATGATGAA 516
QY 1726 CCATTGGATTCCAACTCAGGAGAAAGATGACCTCAATCTTGTCGCTCTTTGGGGTT 1785
DB 517 GCTAAGAAACGCGAGAGAGGCAAAACTAAATTAATCTGTTGAGCAATGTTAGTT 576
QY 1786 CCAGAGATCTGCTTCATCTGAAAGCCCAAGGACCAAGAACTTCAGATCAGACTAGC 1845
DB 577 CCGAGCAGAGCAGCTGGCTGAGACTAAGAAATAATCAGAGAAAGCTTAACAAAGCA 636
QY 1846 ACTGAGAGTGTACCAATGAAAATAACACCA 1877
DB 637 CCAGAACTTACTAAAAAAGCTAAGAGAGCTAA 668

RESULT 13
US-08-247-491A-1
; Sequence 1, Application US/08247491A
; Patent No. 5965400
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: YOTHER, Janet L.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10

```

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
 ; STREET: 745 Fifth Avenue
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10151
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/247,491A
 ; FILING DATE: 23-JUN-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KOWALSKI, Thomas J.
 ; REGISTRATION NUMBER: 32,147
 ; REFERENCE/DOCKET NUMBER: 454312-2041
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-588-0800
 ; TELEFAX: 212-588-0500
 ;
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2085 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; STRAIN: Streptococcus pneumoniae Rx1
 ; IMMEDIATE SOURCE:
 ; CLONE: JY2008
 ; FEATURE:
 ; NAME/KEY: Intron
 ; LOCATION: 1..2085
 ;
 ; NAME/KEY: CDS
 ; LOCATION: Join(127..1983, 1987..1992, 1996..2000
 ; LOCATION: ..2025, 2029..2031, 2035..2085)
 ;
 ; US-08-247-491A-1

Patent No. 5980909
GENERAL INFORMATION:
APPLICANT: Biles, David E.
APPLICANT: Yother, Janet L.
APPLICANT: McDaniel, Larry S
TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface
TITLE OF INVENTION: Protein A
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheomaker and Mattare, Ltd.
STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,795
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,636
FILING DATE: 20-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/048,896
FILING DATE: 20-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/835,698
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/656,773
FILING DATE: 15-FEB-1991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: Rx1
IMMEDIATE SOURCE:
CLONE: JY4313
FEATURE:
NAME/KEY: intron
LOCATION: 1..2085
FEATURE:
NAME/KEY: CDS
LOCATION: join(127..1984)
IS-08-319-795-1

517 GCTAGAAACGCCAAGAAGAGAGGCAAAACTAAATTAACTGTTTCGAGCAATGGTAGTT 576

QY 1851 GAGTGCTACCAATGAAATAACACCAATCTTGACCTCAGTTCCTCAACAGAAAGCCACTGG 1910

[illegible]

Db 511 TGAGGCTCAAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAGAAGAGCTAA 570
Qy 1911 GCCTTCAGCTCATGAAGAA 1929
Db 571 ACAGAAAGCAGCAGAGAA 589

RESULT 19
US-08-961-527-134
; Sequence 134, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-961-527-134
Query Match 1.4%; Score 37.4; DB 4; Length 12665;
Best Local Similarity 44.8%; Pred. No. 1.5;
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 1611 AAAAGCTCCTGAGAAATCATCAGAGGATGTGACAAAATATCAGGAGGAGTATCTGCAGA 1670
Db 1560 AAAAAGTAGCAGAGCTGAGAGAAGGTTGAGAGAGCTAGAAAAGCGGAGATCA 1619

Qy 1671 AAACCCAGTTGAGACCATATCAATATATACACAAATCAGATAGTTTCACAGCCAGCCATT 1730
Db 1620 AAAAGAAGAGATCGCGTAACTACCCCAACCAATACTTACAAAAGCGTTGAACTTGAAT 1679

Qy 1731 GGATTCACACTCAGAGAGAAGATGACCTCAATCTTATCGCTCTTGGGTTCCAGA 1790
Db 1680 TGCTGAGTCCGATGTGGAAGTTAAAAAAGCGGAGCTTGAACCTAGTAAAAAGAGAGCTAA 1739

Qy 1791 AGAATCTGCTTCATCTGAAAAAGCCAGGAACCAAGAACTTCAGATCAGCTAGCACTGA 1850
Db 1740 GGAACCTGAAACGAGGAAAAGATTACACGAAACGGAAGTTGAGATTAAGG 1799

Qy 1851 GAGTGTCTACCAATGAAAATAACACCAATCTCTGAGCTTCAGTTCCAAACAGAGCCAGCTGG 1910
Db 1800 TGAGGCTACAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAGAAGAGCTAA 1859

Qy 1911 GCCTTCAGCTCATGAAGAA 1929
Db 1860 ACAGAAAGCAGCAGAGAA 1878

RESULT 20
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G

US-09-007-005-17
Query Match 1.4%; Score 35.8; DB 4; Length 289;
Best Local Similarity 5.5%; Pred. No. 0.39; Mismatches 101; Indels 0; Gaps 0;
Matches 13; Conservative 101; Mismatches 121; Indels 0; Gaps 0;

Qy 911 AACCCAGTAAAGCGTTTGAGACTTCCTGGTGAATGTCGATGTCAGATAGTGCACCAAGCAAA 970
Db 8 ARCARARURURARURURARURARURARURARURARURARURARURARURARURARURAR 67

Qy 971 GCGCGGAGTGACGAGAAAGAGATGGAGAGCAGAGTCCCAATGTGTCATGTATGCAGA 1030
Db 68 SRNR 127

Qy 1031 GAATCTCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAGGTGCACAAAGCAATA 1090
Db 128 SRNR 187

Qy 1091 GAGGACGAGGAGATCTCGACCCAGAGGTGGAAAGTCAATCAGATATTTCAAC 1145
Db 188 SENNR 242

RESULT 21
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491

RESULT 22
US-09-308-022-5
; Sequence 5, Application US/09308022
; Patent No. 6291654
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
; TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
; TITLE OF INVENTION: PNEUMONIAE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
; STREET: 119 No. 6291654th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/308,022
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20586
; FILING DATE: 12-NOV-1997
; APPLICATION NUMBER: 60/029,444
; FILING DATE: 12-NOV-1996
; APPLICATION NUMBER: 60/038,086
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: 60/059,368
; FILING DATE: 19-SEP-1997
; APPLICATION NUMBER: 60/062,473

```

RECORD 23
US-09-385-982-267
; Sequence 267, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 267
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(598)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-267

```

Query Match	1.48;	Score 35.4;	DB 4;	Length 598;
Best Local Similarity	50.98;	Pred. NO. 0.83;		

THE UNIVERSITY OF CHICAGO

RESULT 26
US-08-299-074A-1/c
; Sequence 1, Application US/08299074A
; Patent No. 5955263
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Michael
; APPLICANT: Sherman, Michael
; TITLE OF INVENTION: SEQUENCE SPECIFIC DNA BINDING
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,074A
; FILING DATE: 01-SEP-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/860,758
; FILING DATE: 31-MAR-1992
; APPLICATION NUMBER: 07/715,182
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32141
; REFERENCE/DOCKET NUMBER: 01107.47071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-299-074A-1

Query Match 1.3%; Score 34.6; DB 2; Length 405;
Best Local Similarity 49.7%; Pred. No. 1.1;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1588 CCACCAGAGGAGACATCTGCAAGAACCCAGTTGAGAACCATATCAATATACACATCA 1647
Db 263 CAACAAGAGTGAACCTGCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204
QY 1648 TATCAGAGAGGAGTATCTGCAAGAACCCAGTTGAGAACCATATCAATATACACATCA 1707
Db 203 GAAGAAG 144
QY 1708 GATAAGTTCACAGCCAGCAATGGATTCCCACTCAGGAGAGAGAGAGAGAGAGAGAGAG 1764
Db 143 AGAAG 87

RESULT 27
US-09-399-773-1/c
; Sequence 1, Application US/09399773
; Patent No. 6245515
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert

; APPLICANT: Kinzler, Kenneth
; APPLICANT: Sherman, Michael
; TITLE OF INVENTION: SEQUENCE SPECIFIC DNA BINDING
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/399,773
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/299,074
; FILING DATE:
; APPLICATION NUMBER: 07/715,182
; FILING DATE: 14-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32141
; REFERENCE/DOCKET NUMBER: 01107.47071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-399-773-1

Query Match 1.3%; Score 34.6; DB 4; Length 405;
Best Local Similarity 49.7%; Pred. No. 1.1;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 1588 CCACCAGAGGAGACATCTGCAAGAACCCAGTTGAGAACCATATCAATATACACATCA 1647
Db 263 CAACAAGAGTGAACCTGCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204
QY 1648 TATCAGAGAGGAGTATCTGCAAGAACCCAGTTGAGAACCATATCAATATACACATCA 1707
Db 203 GAAGAAG 144
QY 1708 GATAAGTTCACAGCCAGCAATGGATTCCCACTCAGGAGAGAGAGAGAGAGAGAGAGAG 1764
Db 143 AGAAG 87

RESULT 28
US-08-606-288-6/c
; Sequence 6, Application US/08606288
; Patent No. 5955087
; GENERAL INFORMATION:
; APPLICANT: Whittle, N.R.
; APPLICANT: Carmichael, J.P.
; APPLICANT: Connor, S.E.
; APPLICANT: Thompson, H.S.G.
; APPLICANT: Wilson, M.J.
; TITLE OF INVENTION: Polypeptides useful as immunotherapeutic
; TITLE OF INVENTION: Agents, and Methods of Polypeptide Preparation
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Suite 3400, Four Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/606,288
FILING DATE: 23-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9503786.7
FILING DATE: 24-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000034
FILING DATE: 08-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9515478.7
FILING DATE: 28-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Walter H. Dreger
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-63284/WHd
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1815 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 7..1812
US-08-606-288-6

Query Match 1.3%; Score 34.6; DB 2; Length 1815;
Best Local Similarity 50.3%; Pred. No. 3;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 1588 CCACGAGGAGGAGCAGTGAACAAAGCTCTGAAGAAATCATCAGAGGATGTGACAAA 1647
DB 350 CCACGTAGTAATAGAGGTTTGCAGAGTTCTTAAGGGAACATAGCCAGTACGCCCA 291
QY 1648 TATCAGGAGGAGTATCTGCAGAAACCCAGTTGAGAACCATATCAATATACAAATCA 1707
DB 290 GTGCGGGAACCGGTGCCTATACCAACCTCGGAAGACACCCCAACTCCCATTTT 231
QY 1708 GATAGTTTCAGAGCAACCCATTGGATTCCTCACTCAGGAGGAAGATG 1756
DB 230 AATATTGATCTGCAATGGTGTGTGTTCCACCTTAGGAATTACATCTG 182

RESULT 29
US-09-347-483-6/C
; Sequence 6, Application US/09347483
; Patent No. 6123948
; GENERAL INFORMATION:
; APPLICANT: Whittle, N.R.
; APPLICANT: Carmichael, J.P.
; APPLICANT: Connor, S.E.
; APPLICANT: Thompson, H.S.G.
; APPLICANT: Wilson, W.J.
; TITLE OF INVENTION: Polypeptides Useful as Immunotherapeutic
; TITLE OF INVENTION: Agents, and Methods of Polypeptide Preparation
; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Suite 3400, Four Embarcadero Center
CITY: San Francisco
STATE: California
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/347,483
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/606,288
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000034
FILING DATE: 08-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9515478.7
FILING DATE: 28-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Walter H. Dreger
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-63284/WHd
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1815 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 7..1812
US-09-347-483-6

Query Match 1.3%; Score 34.6; DB 3; Length 1815;
Best Local Similarity 50.3%; Pred. No. 3;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 1588 CCACGAGGAGGAGCAGTGAACAAAGCTCTGAAGAAATCATCAGAGGATGTGACAAA 1647
DB 350 CCACGTAGTAATAGAGGTTTGCAGAGTTCTTAAGGGAACATAGCCAGTACGCCCA 291
QY 1648 TATCAGGAGGAGTATCTGCAGAAACCCAGTTGAGAACCATATCAATATACAAATCA 1707
DB 290 GTGCGGGAACCGGTGCCTATACCAACCTCGGAAGACACCCCAACTCCCATTTT 231
QY 1708 GATAGTTTCAGAGCAACCCATTGGATTCCTCACTCAGGAGGAAGATG 1756
DB 230 AATATTGATCTGCAATGGTGTGTGTTCCACCTTAGGAATTACATCTG 182

RESULT 30
US-09-007-005-32
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B

; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
US-09-007-005-32

Query Match 1.3%; Score 34.4; DB 4; Length 248;
Best Local Similarity 20.9%; Pred. No. 0.95;
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;
QY 847 AAAGATGATACACGAGCACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906
Db 67 RARGRARCRU 126
QY 907 CGACAACACCAGCTTAAGCGTTTGAAGACTTCGTGGTGGTTCAGATCTGGACCCAGA 966
Db 127 RRCRURGRURARARARARARARARARARARARARARARARARARARARCR 186
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGT 1018
Db 187 RCARCRU 238

RESULT 31
US-09-244-796-32
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 248
; TYPE: RNA
; ORGANISM: Homo sapiens
US-09-244-796-32

Query Match 1.3%; Score 34.4; DB 4; Length 248;
Best Local Similarity 20.9%; Pred. No. 0.95;
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;
QY 847 AAAGATGATACACGAGCACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906
Db 67 RARGRARCRU 126
QY 907 CGACAACACCAGCTTAAGCGTTTGAAGACTTCGTGGTGGTTCAGATCTGGACCCAGA 966
Db 127 RRCRURGRURARARARARARARARARARARARARARARARARARARARCR 186
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGT 1018
Db 187 RCARCRU 238

RESULT 32
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-007-005-3

Query Match 1.3%; Score 34.4; DB 4; Length 277;
Best Local Similarity 20.9%; Pred. No. 1;
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;
QY 847 AAAGATGATACACGAGCACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906
Db 67 RARGRARCRU 126
QY 907 CGACAACACCAGCTTAAGCGTTTGAAGACTTCGTGGTGGTTCAGATCTGGACCCAGA 966
Db 127 RRCRURGRURARARARARARARARARARARARARARARARARARARARCR 186
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGT 1018
Db 187 RCARCRU 238

RESULT 33
US-09-244-796-3
; Sequence 3, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 277
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-244-796-3


```

; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
US-07-885-970A-3

```

Query Match 1.3%; Score 34.4; DB 1; Length 913;
Best Local Similarity 46.9%;
Pred. NO. 2.2;
Matches 107; Conservative 0; Mismatches 121; Indels

		1105	TCTCGACCCAGAGTGGAACAAAGTCAAATCAGATTATTTCAACTTCTCTACGGTCCCACATCA	1164
QY				
		297	TCTCCACCACCAGTTTTCAACTCTCCACCCAGTTCTCTGCTGTGGAATCCACCACT	356
Db				
		1165	AGTCCTGATTTGGAAAGTGAGTGAACCTGCCAATGGAAGTAGATACTCCAGCTGGAACAATTT	1224
QY				
		357	GCTTCTCCTCCTCGTAACCTCCACTCCAGCTTCTCCACTCTCGCACTCTCTCCACCA	416
Db				
		1225	CATTACGCCTTCTACATCCTCTACAATTCAGCTCAGGCTCATTCGCAGCATCATCTCCACACA	1284
QY				
		417	GCPTCTCCACCTCCGSCACTCACCAACCTGCAACCCACCGCCAGCAACTCTCTCTCTCT	476
Db				
		1285	GAAAGCGCTCATTTACTCTCTTTGCTATCTTCTCCACAGAGTGAACAA	1332
QY				
		477	GCTPACCCACACCAAGCTGCATTGGCTTCTCTCCACGCCACAGTCCCA	524
Db				

RESULT 36

US-08-298-687A-3
; Sequence 3, Application US/08298687A
; Patent No. 5521078

; GENERAL INFORMATION:

: APPLICANT: John, Malivakal E.
 : TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
 : TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
 : NUMBER OF SEQUENCES: 33
 : CORRESPONDENCE ADDRESS:
 : ADDRESSER: Nicholas J. Seay, Quarles & Brady
 : STREET: P.O. Box 2113, First Wisconsin Plaza
 : CITY: Madison
 : STATE: Wisconsin
 : COUNTRY: USA
 : ZIP: 53701

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
;
```

CLASSIFICATION: 800
PRIOR APPLICATION DATA: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA: US 07/253,243
FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 3:

```

; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 913 base pairs
; ; TYPE: nucleic acid
; ; STRANDEDNESS: double
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: cDNA
; ; HYPOTHETICAL: NO
; ; ANTI-SENSE: NO

```

ORIGINAL SOURCE:

ORGANISM: Cossypium hirsutum
 STRAIN: Coker 312
 DEVELOPMENTAL STAGE: 15 day old fiber cells
 TISSUE TYPE: fiber cells
 IMMEDIATE SOURCE:
 LIBRARY: CKFBI51
 CLONE: H6
 US-08-298-687A-3

```
Query Match      1.3%; Score 34.4; DB 1; Length 913;
Best Local Similarity 46.9%; Pred. No. 2.2;
Matches 107; Conservative 0; Mismatches 121; Indels
```

[illegible]

RESULT 37

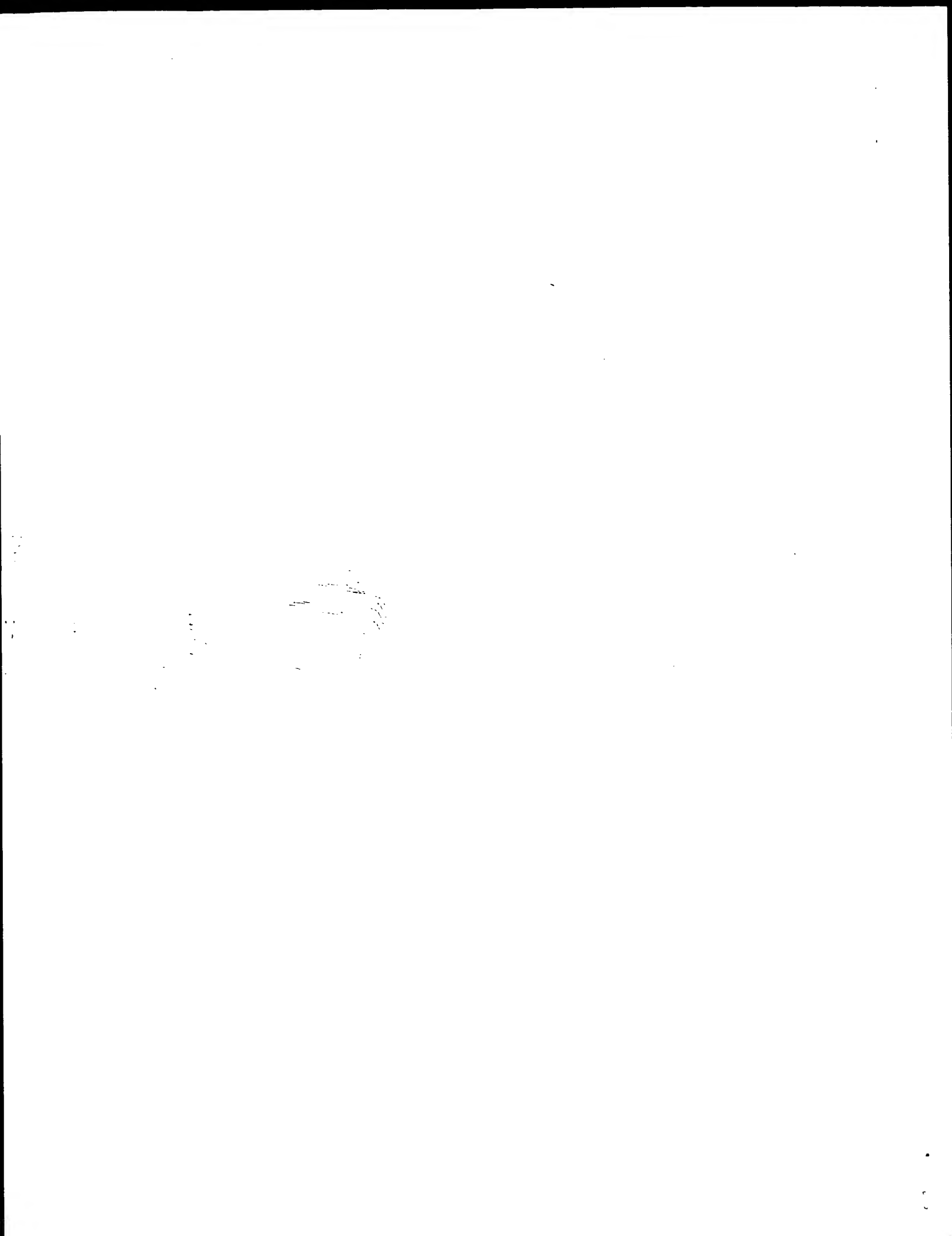
US-08-530-797-2
; Sequence 2, Application US/08530797
; Patent No. 5597718
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O BOX 2113
; STREET: FIRST WISCONSIN PLAZA
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; Zip: 53701

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/617,239
FILING DATE: 21-NOV-90
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-88
ATTORNEY/AGENT INFORMATION:
NAME: Nicholas J. Seay
REGISTRATION NUMBER: 27,386
REFERENCE/SOCKET NUMBER: 1122990245
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 base pairs
TYPE: nucleic acid
STRANDEDNESS: single



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2003, 02:00:09 ; Search time 315 Seconds
(without alignments)
6969.874 Million cell updates/sec

Title: US-09-781-693A-1
Perfect score: 2580
Sequence: 1 atgtctcggtggtgctctc.....atgaaatgaggtaggaa 2580

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database : Published Applications, NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2580	100.0	2580	9	US-09-781-693A-1
2	2580	100.0	3016	9	US-09-781-693A-3
3	1542.4	59.8	1942	9	US-09-822-846-200
4	328	12.7	332	10	US-09-864-761-19286
5	261	10.1	261	10	US-09-864-761-17295
6	255.8	9.9	481	10	US-09-864-761-2560
7	241.8	9.4	409	10	US-09-983-965-58
8	186	7.2	186	10	US-09-864-761-17296
9	160	6.2	467	10	US-09-864-761-499
10	141	5.5	468	10	US-09-864-761-500
11	71.2	2.8	356	10	US-09-770-791-605
12	59	2.3	257	10	US-09-983-965-46
13	54.6	2.1	2679	10	US-09-925-300-645
14	41	1.6	2276	10	US-09-828-447-6
15	39.2	1.5	640661	10	US-09-790-988-1
16	38.6	1.5	1945	10	US-09-925-300-145
17	37.4	1.4	1360	10	US-09-765-272-37
18	37.4	1.4	2794	10	US-09-917-800A-1663
19	36.6	1.4	574	9	US-10-184-644-234

36.6	1.4	671	9	US-10-184-644-346	Sequence 346, App
36.2	1.4	276	10	US-09-864-761-20595	Sequence 20595, A
36.2	1.4	295	10	US-09-823-876-4276	Sequence 4276, Ap
36.2	1.4	462	10	US-09-864-761-3829	Sequence 3829, Ap
35.6	1.4	561	10	US-09-998-598-1297	Sequence 1297, Ap
35.4	1.4	528	9	US-10-152-661-618	Sequence 618, App
35.4	1.4	528	9	US-09-866-050A-618	Sequence 618, App
35.4	1.4	594	10	US-09-925-301-690	Sequence 690, App
35.4	1.4	895	10	US-09-770-445-480	Sequence 480, App
35.4	1.4	1056	12	US-10-044-090-8	Sequence 8, Appli
35.4	1.4	1188	10	US-09-998-598-368	Sequence 368, App
35.4	1.4	1548	10	US-09-998-598-357	Sequence 357, App
35.2	1.4	3972	10	US-09-801-368-33	Sequence 33, Appl
35	1.4	429	10	US-09-860-432-73	Sequence 73, Appl
35	1.4	453	10	US-09-998-598-61	Sequence 61, Appl
35	1.4	559	10	US-09-864-761-7684	Sequence 7684, Ap
35	1.4	4301	9	US-09-989-919-69	Sequence 69, Appl
35	1.4	6823	9	US-09-989-920-16	Sequence 16, Appl
34.8	1.3	3929	10	US-09-867-550-1267	Sequence 1267, Ap
34.8	1.3	3953	10	US-09-864-761-19041	Sequence 19041, A
34.8	1.3	7287	10	US-09-070-927A-210	Sequence 210, App
34.6	1.3	405	9	US-09-813-824A-1	Sequence 1, Appli
34.6	1.3	2000	9	US-09-938-842A-4180	Sequence 4180, Ap
34.6	1.3	11802	7	US-08-781-986A-70	Sequence 70, Appl
34.4	1.3	309	10	US-09-764-887-368	Sequence 368, App
34.4	1.3	309	10	US-09-764-887-369	Sequence 369, App
34.4	1.3	309	10	US-09-764-887-370	Sequence 370, App
34.4	1.3	589	10	US-09-864-761-6977	Sequence 6977, Ap
34.4	1.3	701	10	US-09-864-761-19757	Sequence 19757, A
34.4	1.3	701	10	US-09-864-761-26789	Sequence 26789, A
34.4	1.3	963	10	US-09-864-761-10148	Sequence 10148, A
34.4	1.3	1336	10	US-09-799-777-114	Sequence 114, App
34.4	1.3	1341	10	US-09-814-122-14	Sequence 14, Appl
34.4	1.3	1646	10	US-09-816-828-3	Sequence 3, Appli
34.4	1.3	1934	10	US-09-864-761-2976	Sequence 2976, Ap
34.4	1.3	64667	9	US-10-274-409-3	Sequence 3, Appli
34.2	1.3	848	10	US-09-770-445-632	Sequence 632, App
34.2	1.3	4437	10	US-09-801-368-51	Sequence 51, Appl
34	1.3	320	10	US-09-764-887-40	Sequence 40, Appl
34	1.3	503	9	US-09-981-876-19	Sequence 19, Appl
34	1.3	503	9	US-09-148-545-19	Sequence 19, Appl
33.8	1.3	420	10	US-09-864-761-5017	Sequence 5017, Ap
33.8	1.3	452	10	US-09-864-761-10129	Sequence 10129, A
33.8	1.3	458	9	US-10-184-644-546	Sequence 546, App
33.8	1.3	811	7	US-10-184-644-414	Sequence 414, App
33.8	1.3	13321	7	US-08-781-986A-4	Sequence 4, Appli
33.6	1.3	189	10	US-09-864-761-23708	Sequence 23708, A
33.6	1.3	499	9	US-10-040-739-194	Sequence 194, App
33.6	1.3	526	10	US-09-864-761-7895	Sequence 7895, Ap
33.6	1.3	2000	9	US-09-938-842A-3975	Sequence 3975, Ap
33.6	1.3	4656	10	US-09-864-761-19341	Sequence 19341, A
33.6	1.3	6233	12	US-10-044-090-371	Sequence 371, App
33.6	1.3	7100	10	US-09-932-183A-1	Sequence 1, Appli
33.6	1.3	465327	10	US-09-933-267A-1	Sequence 1, Appli
33.4	1.3	497	10	US-09-783-590-8295	Sequence 8295, Ap
33.4	1.3	734	9	US-10-184-644-458	Sequence 458, App
33.4	1.3	3458	7	US-08-781-986A-317	Sequence 317, App
33.2	1.3	2852	10	US-09-922-261-205	Sequence 205, App
33.2	1.3	306	10	US-09-922-261-203	Sequence 203, App
33.2	1.3	696	10	US-09-922-261-193	Sequence 193, App
33.2	1.3	699	10	US-09-922-261-191	Sequence 191, App
33.2	1.3	700	9	US-10-040-739-1385	Sequence 1385, Ap
33.2	1.3	717	10	US-09-922-261-189	Sequence 189, App
33.2	1.3	774	10	US-09-922-261-187	Sequence 187, App
33.2	1.3	819	10	US-09-922-261-185	Sequence 185, App
33.2	1.3	957	10	US-09-735-705-113	Sequence 113, App
33.2	1.3	957	10	US-09-850-710A-113	Sequence 113, App
33.2	1.3	957	10	US-09-897-778-113	Sequence 113, App
33.2	1.3	1669	10	US-09-922-261-184	Sequence 184, App
33.2	1.3	1915	9	US-09-938-842A-3951	Sequence 3951, Ap
33.2	1.3	2870	9	US-10-174-590-473	Sequence 473, App
33.2	1.3	2870	9	US-10-176-758-473	Sequence 473, App
33.2	1.3	2870	9	US-10-175-737-473	Sequence 473, App

C 93 33.2 1.3 2870 9 US-10-173-706-473 Sequence 473, App
C 94 33.2 1.3 2870 9 US-10-175-738-473 Sequence 473, App
C 95 33.2 1.3 2870 9 US-10-175-752-473 Sequence 473, App
C 96 33.2 1.3 2870 9 US-10-176-482-473 Sequence 473, App
C 97 33.2 1.3 2870 9 US-10-176-757-473 Sequence 473, App
C 98 33.2 1.3 2870 9 US-10-176-913-473 Sequence 473, App
C 99 33.2 1.3 2870 9 US-10-180-552-473 Sequence 473, App
C 100 33.2 1.3 2870 9 US-10-180-557-473 Sequence 473, App
C 101 33.2 1.3 2870 9 US-10-173-700-473 Sequence 473, App
C 102 33.2 1.3 2870 9 US-10-173-702-473 Sequence 473, App
C 103 33.2 1.3 2870 9 US-10-174-572-473 Sequence 473, App
C 104 33.2 1.3 2870 9 US-10-174-579-473 Sequence 473, App
C 105 33.2 1.3 2870 9 US-10-174-582-473 Sequence 473, App
C 106 33.2 1.3 2870 9 US-10-174-588-473 Sequence 473, App
C 107 33.2 1.3 2870 9 US-10-175-730-473 Sequence 473, App
C 108 33.2 1.3 2870 9 US-10-175-740-473 Sequence 473, App
C 109 33.2 1.3 2870 9 US-10-175-743-473 Sequence 473, App
C 110 33.2 1.3 2870 9 US-10-176-488-473 Sequence 473, App
C 111 33.2 1.3 2870 9 US-10-176-492-473 Sequence 473, App
C 112 33.2 1.3 2870 9 US-10-176-747-473 Sequence 473, App
C 113 33.2 1.3 2870 9 US-10-176-750-473 Sequence 473, App
C 114 33.2 1.3 2870 9 US-10-176-985-473 Sequence 473, App
C 115 33.2 1.3 2870 9 US-10-176-991-473 Sequence 473, App
C 116 33.2 1.3 2870 9 US-10-176-992-473 Sequence 473, App
C 117 33.2 1.3 2870 9 US-10-176-993-473 Sequence 473, App
C 118 33.2 1.3 2870 9 US-10-184-658-473 Sequence 473, App
C 119 33.2 1.3 2870 9 US-10-227-884-219 Sequence 219, App
C 120 33.2 1.3 2870 9 US-10-173-695-473 Sequence 473, App

ALIGNMENTS

RESULT 1

US-09-781-693a-1

; Sequence 1, Application US/09781693a

; Publication No. US20030054438A1

; GENERAL INFORMATION:

; APPLICANT: Chang, Tai-Jay

; TITLE OF INVENTION: ANDROGEN RECEPTOR COMPLEX-ASSOCIATED

; FILE REFERENCE: 11709-003001

; CURRENT APPLICATION NUMBER: US/09/781.693A

; PRIOR FILING DATE: 2002-07-23

; PRIOR FILING DATE: 2001-01-17

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2580

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-781-693a-1

Query Match 100.0%; Score 2580; DB 9; Length 2580;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTCGGGTGGCTCTACCCACACCTGTGTGGGACGTGAGGAAAAGGTCCCTCGGG 60
DB 1 ATGCTCGGGTGGCTCTACCCACACCTGTGTGGGACGTGAGGAAAAGGTCCCTCGGG 60
QY 61 CTGGAGGACCGTCCCGGCTCGGAGTGCCTACCTGGGAAGAGAAATTTATCAAAAGA 120
DB 61 CTGGAGGACCGTCCCGGCTCGGAGTGCCTACCTGGGAAGAGAAATTTATCAAAAGA 120
QY 121 TTAACACTGAGCAACCTTAATGTCATGATGCTGTGTTAATACAACTCTGTTGGAA 180
DB 121 TTAACACTGAGCAACCTTAATGTCATGATGCTGTGTTAATACAACTCTGTTGGAA 180
QY 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCT 240
DB 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCT 240

QY 241 TACAGCAGAAAGGTTTTGACAACAATTCGTTAGGGCACCAGCAACATATATTTAGTGCA 300
DB 241 TACAGCAGAAAGGTTTTGACAACAATTCGTTAGGGCACCAGCAACATATATTTAGTGCA 300
QY 301 AAGTTCCTTACCTTGTACAAATGATAACAGATTGTATCTCTCTGGAGATGGAGTAATA 360
DB 301 AAGTTCCTTACCTTGTACAAATGATAACAGATTGTATCTCTCTGGAGATGGAGTAATA 360
QY 361 TTTTATACCACGTTGAGCAGATGCGAGAAACCAACAGCAATGCCAATTTACGTGTAT 420
DB 361 TTTTATACCACGTTGAGCAGATGCGAGAAACCAACAGCAATGCCAATTTACGTGTAT 420
QY 421 TATGGAATCTACTTATGAGATTATGACTGTATACCAATGACCCCTTACACCTTCTCTCTGT 480
DB 421 TATGGAATCTACTTATGAGATTATGACTGTATACCAATGACCCCTTACACCTTCTCTCTGT 480
QY 481 GGTGAAGTGAACCTGTAGGTGTTGATACAGCATCAAACTAGCTGCACAAAGAA 540
DB 481 GGTGAAGTGAACCTGTAGGTGTTGATACAGCATCAAACTAGCTGCACAAAGAA 540
QY 541 GATTGTAAGATGATATTTTAACTGTCGACGTGTCGACGTCTGTGCTATTTTC 600
DB 541 GATTGTAAGATGATATTTTAACTGTCGACGTGTCGACGTCTGTGCTATTTTC 600
QY 601 CCACCAATACCATATTTACCTTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
DB 601 CCACCAATACCATATTTACCTTGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
QY 661 CGGCAATGCTGGCACAAGAGCTACAGGAATATGAGGTGAGGAGGAGGAGGAGGAGGAGGAG 720
DB 661 CGGCAATGCTGGCACAAGAGCTACAGGAATATGAGGTGAGGAGGAGGAGGAGGAGGAGGAG 720
QY 721 GTTGCCCTTTTATTCCTCCCATCTTAATAAGTCTCTGACAGTGCACATCTCTGTGT 780
DB 721 GTTGCCCTTTTATTCCTCCCATCTTAATAAGTCTCTGACAGTGCACATCTCTGTGT 780
QY 781 TACAGTGAAGATGTTCAAGAGATTCCTGTTAGTTACTTCTGATTACATATATCTTTT 840
DB 781 TACAGTGAAGATGTTCAAGAGATTCCTGTTAGTTACTTCTGATTACATATATCTTTT 840
QY 841 GACCGAAAGATGATACAGCAGGAGAACTTAAACTCTCTGCGGAAAGAGAGAGAGAA 900
DB 841 GACCGAAAGATGATACAGCAGGAGAACTTAAACTCTCTGCGGAAAGAGAGAGAGAA 900
QY 901 GAGTTGCGACACCAACCAGTTAAGCGTTTGAAGCTTCGTGTTGTTGTTGTTGTTGTTGTTG 960
DB 901 GAGTTGCGACACCAACCAGTTAAGCGTTTGAAGCTTCGTGTTGTTGTTGTTGTTGTTGTTG 960
QY 961 CCCAGAGCAAGCCGGAGAGTGAACGAGAGATGAGAGCAGATGCCAATCTGTCA 1020
DB 961 CCCAGAGCAAGCCGGAGAGTGAACGAGAGATGAGAGCAGATGCCAATCTGTCA 1020
QY 1021 TTGATGACAGAGATGCTGATATGTTTATCAAGTGGTTTGAAGAGCAAGTGCAGTTGCA 1080
DB 1021 TTGATGACAGAGATGCTGATATGTTTATCAAGTGGTTTGAAGAGCAAGTGCAGTTGCA 1080
QY 1081 CAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGGTGGAAACAGTCAATCATATTT 1140
DB 1081 CAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGGTGGAAACAGTCAATCATATTT 1140
QY 1141 TCAACTCTTCTTACCGTCCCATCAAGTCTGATTGTTGAGTGGAGTGAAGTGAAGTGAAG 1200
DB 1141 TCAACTCTTCTTACCGTCCCATCAAGTCTGATTGTTGAGTGGAGTGAAGTGAAGTGAAG 1200
QY 1201 GTAGTACTCTCAGCTGAACAATTTCTTACAGCTTCTTACATCTCTTACATGTCAGTCTCAG 1260
DB 1201 GTAGTACTCTCAGCTGAACAATTTCTTACAGCTTCTTACATCTCTTACATGTCAGTCTCAG 1260
QY 1261 GCTCATGTCATATCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1261 GCTCATGTCATATCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320

QY 1321 GACAGTGAACAAAGGCGCTGTTGAGGATCTGTGACAGCACCACACATCATCAGTCTGAT 1380
Db 1321 GACAGTGAACAAAGGCGCTGTTGAGGATCTGTGACAGCACCACACATCATCAGTCTGAT 1380
QY 1381 AACATAATGAAAGCTGAGCCCAACACAGGACAGGTGAACCACTTTAAGTTTGCAC 1440
Db 1381 AACATAATGAAAGCTGAGCCCAACACAGGACAGGTGAACCACTTTAAGTTTGCAC 1440
QY 1441 TACAGCAGAGGAGAACCACTACAGCACAATAAACTTACAGATGAATGGAGC 1500
Db 1441 TACAGCAGAGGAGAACCACTACAGCACAATAAACTTACAGATGAATGGAGC 1500
QY 1501 AGTATGATCAAGTCTAGAGGAACTGGAGCCATTCGAATCTGAGGGTCAGGAGAA 1560
Db 1501 AGTATGATCAAGTCTAGAGGAACTGGAGCCATTCGAATCTGAGGGTCAGGAGAA 1560
QY 1561 TCTTTGCTCCACAGAGCTCAGTCAACACCAAGAGGAGAGTATCTGCAGAAAAGCTCCT 1620
Db 1561 TCTTTGCTCCACAGAGCTCAGTCAACACCAAGAGGAGAGTATCTGCAGAAAAGCTCCT 1620
QY 1621 GAAGATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAAAGCTCCT 1680
Db 1621 GAAGATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAAAGCTCCT 1680
QY 1681 GAGAACCATATCAATATAACCAATCAGATTAAGTTCACAGCCAGCCATTCGATTCGAAC 1740
Db 1681 GAGAACCATATCAATATAACCAATCAGATTAAGTTCACAGCCAGCCATTCGATTCGAAC 1740
QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAATCTGCT 1800
Db 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAATCTGCT 1800
QY 1801 TCATCTGAAAAGCCAGGAAACCAAGAACTTCAGATCAGACTAGCAGTGTGCTAAC 1860
Db 1801 TCATCTGAAAAGCCAGGAAACCAAGAACTTCAGATCAGACTAGCAGTGTGCTAAC 1860
QY 1861 AATGAAATTAACCAATCTGAGCCCTCAGTTCACAAACAGAGCCACTGGGCCCTCAGCT 1920
Db 1861 AATGAAATTAACCAATCTGAGCCCTCAGTTCACAAACAGAGCCACTGGGCCCTCAGCT 1920
QY 1921 CATGAAGAAATCCACAGGAGTCTGCTCTTCAGGACACAGATGATGATGATGAC 1980
Db 1921 CATGAAGAAATCCACAGGAGTCTGCTCTTCAGGACACAGATGATGATGATGAC 1980
QY 1981 CCAGTCTGATCCAGGTCAGAGTATCGAGCAGGAGCTGTGTATGACGCTCTGCTGTT 2040
Db 1981 CCAGTCTGATCCAGGTCAGAGTATCGAGCAGGAGCTGTGTATGACGCTCTGCTGTT 2040
QY 2041 GCCGTATTCAGGAGTCTTCAGACGGAGAAAGAAAGGAAAGAAATGGAAGAAATTCGAT 2100
Db 2041 GCCGTATTCAGGAGTCTTCAGACGGAGAAAGAAAGGAAAGAAATGGAAGAAATTCGAT 2100
QY 2101 ACTTTGAACATTAAGAGCCGCTAGTAAAAATGTTTATAAGGCCATTCGCAATCCAGG 2160
Db 2101 ACTTTGAACATTAAGAGCCGCTAGTAAAAATGTTTATAAGGCCATTCGCAATCCAGG 2160
QY 2161 ACAATGATAAAGAGCCATTTCTGGGTCCTAACTTTGATGATGATGATGATGATGATGAT 2220
Db 2161 ACAATGATAAAGAGCCATTTCTGGGTCCTAACTTTGATGATGATGATGATGATGATGAT 2220
QY 2221 GGCCACATTTTCTGAGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280
Db 2221 GGCCACATTTTCTGAGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280
QY 2281 AATCATGTGTAACGTCCTGACGACACATCCGTTTGAACCAATTTTACGCTCATCTGGC 2340
Db 2281 AATCATGTGTAACGTCCTGACGACACATCCGTTTGAACCAATTTTACGCTCATCTGGC 2340
QY 2341 ATAGATTATGACATAAGATCTGTCACCATTTAGACAGTCAAGGATTTTAAACCGAAAA 2400
Db 2341 ATAGATTATGACATAAGATCTGTCACCATTTAGACAGTCAAGGATTTTAAACCGAAAA 2400
QY 2401 CTTGCTGATGAGTTTATACTGAAACGAACACTCATGCTGGAAGAAACTAGAAAACACCAT 2460

Db 2401 CTTGCTGATGAGTTTATACTGAAAGCACTCATGCTGGAAGAACTAGAAACACCAT 2460
QY 2461 ACAGTTCCAGCCCTCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520
Db 2461 ACAGTTCCAGCCCTCTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520
QY 2521 CGGTTGGAGGTCACAGATCAGAGGCTCTGCTCAAGAGAATCAAAATCAGATGAGGAA 2580
Db 2521 CGGTTGGAGGTCACAGATCAGAGGCTCTGCTCAAGAGAATCAAAATCAGATGAGGAA 2580

RESULT 2

US-09-781-693A-3
; Sequence 3, Application US/09781693A
; Publication No. US20030054438A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Tai-Jay
; TITLE OF INVENTION: ANDROGEN RECEPTOR COMPLEX-ASSOCIATED
; FILE REFERENCE: 11709-003001
; CURRENT APPLICATION NUMBER: US/09/781,693A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/262,312
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 3016
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)...(2597)
US-09-781-693A-3

Query Match 100.0%; Score 2580; DB 9; Length 3016;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTCTGGGTCGCTCTTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCCTCGG 60
Db 18 ATGCTCTGGGTCGCTCTTACCCACACCTGTGTGGGACGTGAGGAAAGGTCCCTCGG 77
QY 61 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTACCTGGAAGAGAAATTTATCCAAAGA 120
Db 78 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTACCTGGAAGAGAAATTTATCCAAAGA 137
QY 121 TTAACACTTGAAGCAACCTTAAATGTGATGATGTTGTGTTAATACATCTGTGGAAT 180
Db 138 TTAACACTTGAAGCAACCTTAAATGTGATGATGTTGTGTTAATACATCTGTGGAAT 197
QY 181 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATGATCCT 240
Db 198 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATGATCCT 257
QY 241 TACAGCAAGAAAGTTTGACAACTTCGTTACGGGACCCAGCAACATATTTAGTGCA 300
Db 258 TACAGCAAGAAAGTTTGACAACTTCGTTACGGGACCCAGCAACATATTTAGTGCA 317
QY 301 AGTTCTTACCTGTGACAAATGATAATGATGATGATGATGATGATGATGATGATGAT 360
Db 318 AGTTCTTACCTGTGACAAATGATAATGATGATGATGATGATGATGATGATGATGAT 377
QY 361 TTTTATACCAAGTTGAGCAAGATGCAAGACCAACACACATGATGATGATGATGATGAT 420
Db 378 TTTTATACCAAGTTGAGCAAGATGCAAGACCAACACACATGATGATGATGATGATGAT 437
QY 421 TATGGAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Db 438 TATGGAATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
QY 481 GGTGAGATGGAATGTTAGTGGTTTGTATACGGCATCAAACTAGCTGCACAAAAGAA 540

Db 498 GGTGAAGATGGAACCTGTTAGGTGTTTATACACGATCAAACTAGCTGCACAAAGAA 557
QY 541 GATTGTAAGATGATATTTAACTGTCGAGCTGCTGCCAGCTGTTGCTATTGTC 600
Db 558 GATTGTAAGATGATATTTAACTGTCGAGCTGCTGCCAGCTGTTGCTATTGTC 617
QY 601 CCACCAATACCAATACCTTGTGTTGTTGTTCTGACAGCTCAGTACCAATATATGAT 660
Db 618 CCACCAATACCAATACCTTGTGTTGTTGTTCTGACAGCTCAGTACCAATATATGAT 677
QY 661 CGGCGAATGCTGGGACACAGAGCTACAGGAAATATGAGGTCAGGAGCTACTGGAATG 720
Db 678 CGGCGAATGCTGGGACACAGAGCTACAGGAAATATGAGGTCAGGAGCTACTGGAATG 737
QY 721 GTTCCCGCTTTTATCTCCCTCCATCTTAATTAATTAAGTCTGACAGTGCATCTGTGT 780
Db 738 GTTCCCGCTTTTATCTCCCTCCATCTTAATTAATTAAGTCTGACAGTGCATCTGTGT 797
QY 781 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTT 840
Db 798 TACAGTGAAGATGCTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTT 857
QY 841 GACCCGAAGATGATACAGCAGAGAACTTAAACTCTGCTCGGAGAGAGAGAA 900
Db 858 GACCCGAAGATGATACAGCAGAGAACTTAAACTCTGCTCGGAGAGAGAGAA 917
QY 901 GAGTTGGACACACCACTGTTAAGCTTTGAGACTTCTGTTGATTTGTTGATCTGGA 960
Db 918 GAGTTGGACACACCACTGTTAAGCTTTGAGACTTCTGTTGATTTGTTGATCTGGA 977
QY 961 CCCAGAGCAGGCGGAGAGTGAACGAGAGAGATGAGAGAGAGAGTCCCAATGTGCA 1020
Db 978 CCCAGAGCAGGCGGAGAGTGAACGAGAGAGATGAGAGAGAGAGTCCCAATGTGCA 1037
QY 1021 TTGATGACAGAACTGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1080
Db 1038 TTGATGACAGAACTGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1097
QY 1081 CAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAACTACATGATAT 1140
Db 1098 CAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAACTACATGATAT 1157
QY 1141 TCAACTCTCTACGGTCCCATCAAGTCTGTTTGAAGTGAAGTGAAGTGAAGTGAAG 1200
Db 1158 TCAACTCTCTACGGTCCCATCAAGTCTGTTTGAAGTGAAGTGAAGTGAAGTGAAG 1217
QY 1201 GTAGATCTCCAGCTGAACAATTTCTCAGCCTTCTACATCTCTACATCTCTACATCT 1260
Db 1218 GTAGATCTCCAGCTGAACAATTTCTCAGCCTTCTACATCTCTACATCTCTACATCT 1277
QY 1261 GGTCAATCGACATCATCTCCACAGAAAGCCCTCATCTACTCTCTCTCTCTCTCTCT 1320
Db 1278 GGTCAATCGACATCATCTCCACAGAAAGCCCTCATCTACTCTCTCTCTCTCTCTCT 1337
QY 1321 GACAGTGAACAAGCGAGTCTGTTGAGGATCTGACAGGAGTGAACCACTGATCTGAT 1380
Db 1338 GACAGTGAACAAGCGAGTCTGTTGAGGATCTGACAGGAGTGAACCACTGATCTGAT 1397
QY 1381 RACAAATATGAAGTGAAGCCCAACAGGAGAGTGAACCACTGATCTGATCTGATCTGAT 1440
Db 1398 RACAAATATGAAGTGAAGCCCAACAGGAGAGTGAACCACTGATCTGATCTGATCTGAT 1457
QY 1441 TACAGCAGAAAGGAACTACAGCAGCAATAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1500
Db 1458 TACAGCAGAAAGGAACTACAGCAGCAATAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1517
QY 1501 AGTATAGCATCAAGTCTAGAGAAATGGAGCCATTTGAGGAGTGAAGTGAAGTGAAGTGAAG 1560
Db 1518 AGTATAGCATCAAGTCTAGAGAAATGGAGCCATTTGAGGAGTGAAGTGAAGTGAAGTGAAG 1577
QY 1561 TCTTTTCTCCACAGAGCTCAGTGAACCCAGAGAGAGAGAGTGAACAAAGAGTCTCT 1620

Db 1578 TCTTTCTCTCCACAGAGCTCAGTGCACACCCAGAGAGAGACAGTGAACAAAGAGTCTCT 1637
QY 1621 GAAGAAATCATCAGAGAGTGTGACAAAATATCAGGAAGAGTATCTGCAGAAAACCCAGTT 1680
Db 1638 GAAGAAATCATCAGAGAGTGTGACAAAATATCAGGAAGAGTATCTGCAGAAAACCCAGTT 1697
QY 1681 GAGAACCATATCAATATAACAAATCAGATTAAGTTTCAGAGCAAGCCATTTGGATTCCAA 1740
Db 1698 GAGAACCATATCAATATAACAAATCAGATTAAGTTTCAGAGCAAGCCATTTGGATTCCAA 1757
QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTGATCGCTTGTGTTGGGTTCCAGAAATCTCT 1800
Db 1758 TCAGGAGAAAGAAATGACCTCAATCTTGTGATCGCTTGTGTTGGGTTCCAGAAATCTCT 1817
QY 1801 TCATCTGAAAAGCAAGGAAACCAAGAACCTTCAGATCAGACTAGACATGAGAGTGTCTAC 1860
Db 1818 TCATCTGAAAAGCAAGGAAACCAAGAACCTTCAGATCAGACTAGACATGAGAGTGTCTAC 1877
QY 1861 AATGAAATTAACCAATCTCTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCT 1920
Db 1878 AATGAAATTAACCAATCTCTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCT 1937
QY 1921 CATGAAGAAACATCCACAGGAGCTCTCTCTCAGGACACAGATGACAGTGTGATGAC 1980
Db 1938 CATGAAGAAACATCCACAGGAGCTCTCTCTCAGGACACAGATGACAGTGTGATGAC 1997
QY 1981 CCAGTCTCTGATCCCAGGTGCAAGGTATCAGACAGACCTGTTGATGAGAGTGTCTCTGT 2040
Db 1998 CCAGTCTCTGATCCCAGGTGCAAGGTATCAGACAGACCTGTTGATGAGAGTGTCTCTGT 2057
QY 2041 GCCCGTATTCAGGAGTCTTTCAGAGGAGAAAGAAAGAAAGAAATGGAATTTGGAT 2100
Db 2058 GCCCGTATTCAGGAGTCTTTCAGAGGAGAAAGAAAGAAAGAAATGGAATTTGGAT 2117
QY 2101 ACTTTGAACATTAAGAGCCGCTAGTAAATGTTTATAAGGCCATCGCAACTCCAGG 2160
Db 2118 ACTTTGAACATTAAGAGCCGCTAGTAAATGTTTATAAGGCCATCGCAACTCCAGG 2177
QY 2161 ACAATGATTAAGAAAGCAATTTCTGGGCTGCTAACTTTGATGAGTGTGTTCTGACGT 2220
Db 2178 ACAATGATTAAGAAAGCAATTTCTGGGCTGCTAACTTTGATGAGTGTGTTCTGACGT 2237
QY 2221 GGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGTGCTGGAAGCTGAT 2280
Db 2238 GGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGTGCTGGAAGCTGAT 2297
QY 2281 AATCATGTTGTAATCGCTGCGACACATCCGTTTGGCCCAATTTTACCTCATCTGCG 2340
Db 2298 AATCATGTTGTAATCGCTGCGACACATCCGTTTGGCCCAATTTTACCTCATCTGCG 2357
QY 2341 ATAGATTAATGACATAAAGATCTGTCACCATTAAGAGAGTCAAGGATTTTAAACCGAAA 2400
Db 2358 ATAGATTAATGACATAAAGATCTGTCACCATTAAGAGAGTCAAGGATTTTAAACCGAAA 2417
QY 2401 CTGCTGATGAAGTTTAACTCGAAACGAACTCATCTGCTGGAAGAACTAGAAACACCAT 2460
Db 2418 CTGCTGATGAAGTTTAACTCGAAACGAACTCATCTGCTGGAAGAACTAGAAACACCAT 2477
QY 2461 ACAGTTTCCAGCTCTTTTCAATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520
Db 2478 ACAGTTTCCAGCTCTTTTCAATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2537
QY 2521 CGGTTGGAGGAGTGAAGATCAGAGCTCTGCTCAAGAGAAATGAAATGAGGATGAGGAA 2580
Db 2538 CGGTTGGAGGAGTGAAGATCAGAGCTCTGCTCAAGAGAAATGAAATGAGGATGAGGAA 2597

RESULT 3

US-09-822-846-200
; Sequence 200, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth


```

; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing

```

RESULT 7
US-09-983-965-58
; Sequence 58, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding

APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 499
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031287.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.81
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
US-09-864-761-499
Query Match 6.2%; Score 160; DB 10; Length 467;
Best Local Similarity 100.0%; Pred. No. 8.6e-37;
Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1116 AGGTGAACAAAGTCAATCAGATATTTCAACTCTTCCACGCTCCCAATCAAGTCTGATT 1175
DB 308 AGGTGAACAAAGTCAATCAGATATTTCAACTCTTCCACGCTCCCAATCAAGTCTGATT 367
QY 1176 GGAAGTGAAGTGAAGTCAATGGAAGTAGATATCCAGCTGAACAAATTTCTTCAGCCCTC 1235
DB 368 GGAAGTGAAGTGAAGTCAATGGAAGTAGATATCCAGCTGAACAAATTTCTTCAGCCCTC 427

QY 1236 TACATCTCTTACATGTCACTCAGCTCAGGCTCATTCGACATCA 1275
DB 428 TACATCTCTTACATGTCACTCAGCTCAGGCTCATTCGACATCA 467
RESULT 10
US-09-864-761-500
Sequence 500, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 500
LENGTH: 468
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031287.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
US-09-864-761-500

Query Match 5.5%; Score 141; DB 10; Length 468;
Best Local Similarity 100.0%; Pred. No. 3.5e-31;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 AGTTTGTGACAAATTCGTTTCAGGGCAGGAGCAACATATTAGTGCAGAAAGTTCTTAC 310
DB 328 AGTTTGTGACAAATTCGTTTCAGGGCAGGAGCAACATATTAGTGCAGAAAGTTCTTAC 387
QY 311 CTGTGCAATGATAACAGATGATCTCTCGAGATGGAGTAATATTATACCA 370
DB 388 CTGTGCAATGATAACAGATGATCTCTCGAGATGGAGTAATATTATACCA 447
QY 371 ACGTTGAGCAAGATGCAGAA 391
DB 448 ACGTTGAGCAAGATGCAGAA 468

RESULT 11

US-09-770-791-605

; Sequence 605, Application US/09770791

; Patent No. US20020620141

; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Krickler, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE OF INVENTION: thaliana

; CURRENT APPLICATION NUMBER: US/09/770,791

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/178,480

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 605

; LENGTH: 356

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-791-605

Query Match 2.8%; Score 71.2; DB 10; Length 356;
Best Local Similarity 55.7%; Pred. No. 1.2e-10;
Matches 136; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 109 TTTATCCAAAGATTAAACTTGAAGCAACCCCTTAATGTGATGATGTTGTGTTAATACA 168
DB 104 TTGTTCCGAGACTTCTCAGGAGCAAGAAATGGAGGACATCAAGCTGTGTTAATGCT 163
QY 169 ATCTGTGGAATGACACTGGAGATATATTTATCTGGCTCAGATGACACCAATATAGTA 228
DB 164 TTACGCGTGAACCTCTAATGTTCACCTTTTGATATCTGGATCAGATGATTTAAGGATCAAT 223
QY 229 ATTAGTAATCTTACACGAGAAAGTTTTCACAACAATTCGTTCCAGGCGCCGAGCAAC 288

DB 224 ATTTGGAATTACTCCAGTCGGAAGCTTTTGCATCTATATACAGGGCATAGTCCCAAC 283
QY 289 ATATTAGTGCAGAAAGTTCTTACCTTGTACAAATGATAAAGATGATGATCTGCTGGA 348
DB 284 ATCTTCGTGACGAAGTTTGTCCCTGAAACCTCTGACGAGCTTGTGGTATCTGCTGGA 343
QY 349 GATG 352
DB 344 GATG 347

RESULT 12

US-09-983-965-46

; Sequence 46, Application US/09983965

; Patent No. US20020137160A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathalagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 37-21(10297)C

; CURRENT APPLICATION NUMBER: US/09/983,965

; CURRENT FILING DATE: 2001-10-26

; PRIOR APPLICATION NUMBER: US 09/465,231

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: US 60/113,678

; PRIOR FILING DATE: 1998-12-17

; NUMBER OF SEQ ID NOS: 5912

; SEQ ID NO 46

; LENGTH: 257

; TYPE: DNA

; ORGANISM: Bos taurus

; FEATURE:

; OTHER INFORMATION: Clone ID: 06-BOVMS1-017-Q1-E1-B9

US-09-983-965-46

Query Match 2.3%; Score 59; DB 10; Length 257;
Best Local Similarity 60.2%; Pred. No. 3.9e-07;
Matches 115; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

QY 2374 GAAGATCAAGGATTTTAAACCGAAACTTCTGATGAAGTTATACCTCGAAACGAATC 2433
DB 65 GAAGCTCGAAGAGTTTACCTTGGTCCCTCTGATCGGTCATCARGGGCTTCTACAC 124
QY 2434 ATGCTGGAAGAACTAGAAACACACATTACAGTTCCAGCCTCTTTTCATGTTGAGGATGTTG 2493
DB 125 AAGCTGGAAGATACCATTCAGTGGATTTCATTTCCACCCGCTTCACGGGAAGATCTTG 184
QY 2494 GTTTCACCTTAATCATATCCGAGCTGACCGGTTGGAGGGTGACAGATCAGAAAGGCTCTGGT 2553
DB 185 GGGTCACTCGAACAATATCCATTTGAGCGATTGG-CAATTCCAAATTAGGGTCTCGGT 243
QY 2554 CAAGAGAATGA 2564
DB 244 CAGAAGACAGA 254

RESULT 13

US-09-925-300-645

; Sequence 645, Application US/09925300

; Patent No. US20020151681A1

; GENERAL INFORMATION:

; APPLICANT: Craig Rosen,

; APPLICANT: Steve Ruben

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: FA101

; CURRENT APPLICATION NUMBER: US/09/925,300

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05988

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

;;
;; PRIOR FILING DATE: 1999-03-12
;; NUMBER OF SEQ ID NOS: 1890
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 645
;; LENGTH: 2679
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (3)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (21)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (24)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (41)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (124)
;; OTHER INFORMATION: n equals a,t,g, or c
;; NAME/KEY: misc feature
;; LOCATION: (128)
;; OTHER INFORMATION: n equals a,t,g, or c
;;
US-09-925-300-645

Query Match 2.1%; Score 54.6; DB 10; Length 2679;
Best Local Similarity 59.3%; Pred. No. 4.2e-05;
Matches 112; Conservative 0; Mismatches 74; Indels 3; Gaps 1;
QY 2193 TAACCTTTGTAAGTAGTGCTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGC 2252
Db 315 TGAGTTTGTGTGAGCGGTAGTGTGCTGTGGCCACATCTCTCTGGGAGAAATCATCTGTG 374
QY 2253 TGACCATTTGATGCTCTGGAAGCTGATAA---TCATCTGTGTAACCTGCCCTGCAGCCACA 2309
Db 375 CCAGATTATTCAGTTCATGAGGGGGACAGGGAGCGGTGTAAACTGCTGTGAGCCCCA 434
QY 2310 TCCGTTTGACCAATTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGTGTCAAC 2369
Db 435 CCCACAGCTGCTGCTGTGGCAACAGTGGCCCTAGACCATGATGTGAAGATCTGGGCACC 494
QY 2370 ATTAGAGA 2378
Db 495 CACAGCTGA 503

RESULT 14
US-09-828-447-6
; Sequence 6, Application US/09828447
; Patent No. US20020069432A1
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNERT, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; APPLICANT: ISHITANI, MANABU
; TITLE OF INVENTION: SIGNAL TRANSDUCTION STRESS-RELATED PROTEINS AND METHODS
; FILE REFERENCE: 16313-0037
; CURRENT APPLICATION NUMBER: US/09/828,447
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2276
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-447-6

Query Match 1.6%; Score 41; DB 10; Length 2276;
Best Local Similarity 47.8%; Pred. No. 0.39;
Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;
QY 1636 GATGTGCAAAATATCAGGAAGGAGTATCTGCAGAAAACCCAGTTGAGAACCATATCAAT 1695
Db 1050 GATAAGTTGCGAGGAGCAGTCAACTGCTCTGTTAAAGCCCGCTTGAGAAAAAGATTGCA 1109
QY 1696 ATAACAATATCAGATAAGTTTCACAGCCCAAGCCATTGGATTCCCACTCAGGAGAAAGAAAT 1755
Db 1110 GTTCCACCATCAGAGAAGCAAAATCCATTTCCGAAGAGAGGAGCTTGAGTGAAAAAGTT 1169
QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGAAATCTGCTTCATCTGAAAAAGCC 1815
Db 1170 GGAATTTTACGTTGATTTCAGAGGGTGAATCAGCTGATCTGCCCCCTCAGAGTTCCCCC 1229
QY 1816 AAGGAACCAAGAACTTCAGACTCAGACTAGCAGTGTGCTTACCAATGAAAAATACACC 1875
Db 1230 CACGGTAAGAAAGCAACATTGACAGCGGATAGTGAAGAGTGACGATGACGACAATAAGAG 1289
QY 1876 AATCCTGAG 1884
Db 1290 AATCCTGAG 1298

RESULT 15
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 1.5%; Score 39.2; DB 10; Length 640681;
Best Local Similarity 50.0%; Pred. No. 77;
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 1603 AGTCAAAACAAAGCTCCTGGAAGATCATCAGAGTGTGCAAAATATCAGAGGAGATA 1662
Db 198822 AGTCACACTAAACAGAGAGATATATTCGATACCGCTTAAACAATCATCTTCAGATGAG 198881
QY 1663 TCTCAGAAAACCCAGCTTGGAACCATATATACACATCAGATAAGTTTCACAGCC 1722
Db 198882 TATACAGAAATATATTTTAAATAATAACACTGATATGATTTAATAAATACTCATAC 198941
QY 1723 AAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTGTGG 1782
Db 198942 AATGAATCAGATATAAATTTTAGAGAAATCAAGTGAAGAAAGATTAATAATAATGTTGAT 199001
QY 1783 GTTCCAGAAGAACTG 1798
Db 199002 GATACAGACGATATTG 199017

RESULT 16
US-09-925-300-145
; Sequence 145, Application US/09925300

```

; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
;
US-09-765-272-37

Query Match 1.4%; Score 37.4; DB 10; Length 1360;
Best Local Similarity 44.8%; Pred. No. 3.1;
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 1611 AAAAGCTCTGAAGATCATCATCAGAGGATGTGACAAATATATCAGGAAGGAGTATCTGCAGA 1670
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 AAAAAGGTGACGAAGCTCAGAGAAGGTTGAAGAAGCTTAAGAAAAAGCCGAGGATCA 330
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1671 AAACCCAGTTTGAGAACCATATCAATATATACACATCAGATAAGTTTCACGCCAAGCCATT 1730
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 AAAAGAAGAAGATCGCGGTAACTACCCACCAATACTTACAAACACGCTTGAACCTTGAAT 390
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1731 GGATTCCAACTCAGGAGAAAGAATGACCTCAATCTTGTATGCTCTTGTGGGTTCCAGA 1790
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 391 TGTGATCGTCGATGTGGAGTTTAAAGACGGGAGCTTGAAC TAGTAAGAGAGGAAAGCTAA 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1791 AGAATCTGCTTCATCTGAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGCACTGA 1850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 451 GGAACCTCGAAACGAGGAAAAAGTTAAGCAAGCAAGCAAGCGGAAGTTGAGAGTAAAAAAGC 510
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1851 GAGTGCTACCAATGAAAATTAACACCAATCTGTAGCCCTCAGTTCCAAAACAGAAAGCACTGG 1910
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 TGAGGCTACAAAGTTAGAAAAATCAGACAGATCGTAAAAACGAGAAAGAGCTAA 570
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1911 GCCTTCAGCTCATGAGAA 1929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 ACGAAAAGCAGCAGAGAA 589
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 18
US-09-917-800A-1663/c
; Sequence 1663, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13


```

; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1663
; LENGTH: 2794
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession NO
US-09-917-800A-1663

```

[illegible]

RESULT 19
US-10-184-644-234/c
/ Sequence 234, Application US/10184644
/ Publication No. US2003004930A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Jian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin

```

: : TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: : TITLE OF INVENTION: ACIDS ENCODING THE SAME
: : FILE REFERENCE: P3430RIC227
: : CURRENT APPLICATION NUMBER: US/10/184,644
: : CURRENT FILING DATE: 2002-06-28
: : Prior Application removed - See File Wrapper or Palm
: : NUMBER OF SEQ ID NOS: 612
: : SEQ ID NO 234
: : LENGTH: 574
: : TYPE: PRT
: : ORGANISM: Homo Sapien
US-10-184-644-234

```

Query Match	1.4%;	Score 36.6;	DB 9;	Length 574;
Best Local Similarity	7.5%;	Pred. NO. 2.9;		
Matches 36;	Conservative 143;	Mismatches 301;	Indels 0;	Gaps
QY	2	TGTCTCGGGGTGGCTCCACACACCTGTTGTGGAGCTGAGGAAAAGGTCCTCGGGC	61	
	:::	:::	:::	:::
Ddb	496	HSB...SN...S.WHB.MTTBKS.B.TS.BS.SBR.CDN.H.AR.AWM..TY....SBM	437	
QY	62	TGGAGGACCGCTCCCGGCTGCGAGTCGCTACCTGGGAAGAGACAAATTATCCAAAGAT	121	
	:::	:::	:::	:::
Ddb	436	ABSNT.....K...D..YH.T..RD.RS.MN.MCMSA..A.ADBR.Y.RW.BAH...B...	377	
QY	122	TAAAACTTGAAGCAACCCCTTAATGTCATGATGGTTGTGTTAATACAATCTGTTGGAATG	181	

[illegible]

	Query Match	1.4%;	Score 36.6;	DB 9;	Length 671;
	Best Local Similarity	8.9%;	Pred. No. 3.2;	Mismatches 32;	Conservative 139; Indels 189; Gaps 1;
QY	1322	ACAGTGAACAAGCAGCTGCTTGTCAGGCATCTGGACACCACACATCATCAGTCTGATA	1381		
Db	51	AFLGPKDLFFYDKCKDYKRPKKRGNEGLWETQNPHASYSAPPPVSSSDSEAPEANP	110		
QY	1382	ACAAATATGAAGACTGAGCCCCAACCCAGGACAGGTGAACCACTTTTAAGTTTGTCAC	1441		
Db	111	ADGSDADEDDRGMVAVTAVTATAASDRMESDSOSNGLKRTPALKMSVSakra	170		
QY	1442	ACAGCACAGAAGGA - ACAACTTACAAGCACATAAATCGTAACTTTACAGATGAATGAGC	1500		
Db	171	RKASSDLDQASVSPSEENSESSESSEKTSDODTFPEKKAIVAPRRGPLGRKKKKAPS	230		
QY	1501	ACATATACCATCAAGTTCTACAGNAATTGGAGCCATCTGCAAATCTGAGGGTCAGGAGAA	1560		
Db	231	ASDSDSKADSGAKPEPVAMARSASSSSSSSSSSSDSOVKKKPPRGKPAKPLPKPRGR	290		

APPLICANT: HUIDAN, PALLICK
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
TITLE OF INVENTION: thaliana
FILE REFERENCE: 2023US (PARA-012PRV)

— — — — —

Db	395	CAGATCTCGAATTCCAAATGAGAGCAGCCCAATACAGAGAAGCATTCGCACATCTGTTTTC	336
Qy	1238	CATCCTCTACAATGTGAGTCTCAGGCTCATTCGACATCATCTCCCA	1282
Db	335	CACAGTTTTTCATGGCCACAGCAGCCACAGCAGTCATCCTCTGTTTCCA	291

```

RESULT 32
US-09-801-368-33
; Sequence 33, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 3972
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-33

```

RESULT 33
US-09-860-432-73
; Sequence 73, Application US/09860432
; Patent No. US20020081606A1
; GENERAL INFORMATION:
; APPLICANT: TRIEU-CUOT, Patrick
; APPLICANT: POVART, Claire
; TITLE OF INVENTION: METHODS FOR DETECTING AND IDENTIFYING GRAM POSITIVE BACTERIA IN A
; FILE REFERENCE: 2064270S0
; CURRENT APPLICATION NUMBER: US/09/860.432
; CURRENT FILING DATE: 2001-09-12

```

; PRIOR APPLICATION NUMBER: 60/205,237
;
; PRIOR FILING DATE: 2000-05-19
;
; NUMBER OF SEQ ID NOS: 96
;
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 73
;
; LENGTH: 429
;
; TYPE: DNA
;
; ORGANISM: Staphylococcus hyicus
;
US-09-860-432-73

```

```

RESULT 34
US-09-998-598-61/c
: Sequence 61, Application US/09998598
: Patent No. US20020150922A1
: GENERAL INFORMATION:
: APPLICANT: Stolk, John A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Chenault, Ruth A.
: APPLICANT: Mesgher, Madeleine Joy
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE OF INVENTION: DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.561
: CURRENT APPLICATION NUMBER: US/09/998,598
: CURRENT FILING DATE: 2001-11-16
: NUMBER OF SEQ ID NOS: 2606
: SOFTWARE: Corixa Invention Disclosure Database
: SEQ ID NO 61
: LENGTH: 453
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-998-598-61

```

RESULT 35
US-09-864-761-7684/c
; Sequence 7884, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

Query Match 1.4%; Score 35; DB 9; Length 6823;
Best Local Similarity 48.7%; Pred. No. 52;
Matches 95; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1287 AAGCCCTCACTTCTACTCTTCTGCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTGTA 1346
DB 5321 AAACCCCTCAACCTAAACCTGGGTTGGGCTCTAAAAGAAATTTAGGGGAAGCAGATCA 5262
QY 1347 GGCATCTGGACACACACACATCATCTAGTCTGATAACAATAATGAAAGCTGAGCCCAA 1406
DB 5261 AGACCTGATATCAGTCTCAGATGGAATCTCAACAGATTTCAAAAGCAACTCACCAC 5202
QY 1407 ACCAGGACAGGTGAACAGGTTTAAAGTTTGCATCAGCAGACAGGAAGCAACTACAAAG 1466
DB 5201 AAGTGTGAGACACAAAGAACTTTAAATATCATCTAGCAAACTACACTAAGAACTCAATG 5142
QY 1467 CACAATAAACTGAA 1481
DB 5141 ACCAAATAATCA 5127

RESULT 38
US-09-867-550-1267/c
; Sequence 1267, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1267
; LENGTH: 929
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: Wherein n is one of a or t or c or g
US-09-867-550-1267

Query Match 1.3%; Score 34.8; DB 10; Length 929;
Best Local Similarity 45.8%; Pred. No. 14;
Matches 120; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1747 GAAAGAAATGACCTCACTTCTGATGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCT 1806
DB 501 GAAAGCCGAGAGGTCTACCAAGACCCAGGTGGGATCATATAAGAGAGCGGGTGGCAGCC 442
QY 1807 GAAAGAGCCAGGACCAAGCAATTCAGATCAGATAGCTAGAGTGTACCAATGAA 1866
DB 441 CAGGATGTTGAGACGCGGAGATGTTGATCGTGAATCTCGCAAGACCTTTTGGAGAA 382
QY 1867 AATACACCAATCTCTGAGCCCTGCTTCCAAACAGAACCACTGGGCCCTTCAGCTCATGAA 1926
DB 381 AATGTTCTGAAGCAGAGACAGAGCGGTGGCGGCGGCTGAGAGTCCCAATGTC 322
QY 1927 GAAACATCCAGGAGTCTGCTCTTCCAGACACAGATGACAGTGTGATGAGCAGCCAGTC 1986
DB 321 AAAAGACAACCTCGGTGAGCGGTGGCGGACGAGCAGCCGCTGATGATGGGGATGAGTGACAA 262
QY 1987 CTGATCCAGGTGCAAGGTATC 2008

Db 261 GTATACCTTGCTGCTCTCTTC 240

RESULT 39

US-09-864-761-19041/c
; Sequence 19041, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19041
; LENGTH: 3953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP00511.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: AW867076.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 5.00e-63
; OTHER INFORMATION: NT HIT: AE001609.1, EVALUE 6.50e-01
US-09-864-761-19041

[illegible]

